
# Integrated Switch Technology

This section discusses the following topics:

- Cisco Intelligent Gigabit Ethernet Switch Module for the IBM BladeCenter
- Cisco Gigabit Ethernet Switch Module for the HP BladeSystem

# Cisco Intelligent Gigabit Ethernet Switch Module for the IBM BladeCenter

This section provides best design practices for deploying Cisco Intelligent Gigabit Ethernet Switch Modules (Cisco IGESMs) for the IBM eServer BladeCenter (BladeCenter) within the Cisco Data Center Networking Architecture. This section describes the internal structures of the BladeCenter and the Cisco IEGSM and explores various methods of deployment. It includes the following sections:

- Cisco Intelligent Gigabit Ethernet Switching Module
- Cisco IGESM Features
- Using the IBM BladeCenter in the Data Center Architecture
- Design and Implementation Details

## Cisco Intelligent Gigabit Ethernet Switching Module

This section briefly describes the Cisco IGESM and explains how the blade servers within the BladeCenter chassis are physically connected to it.

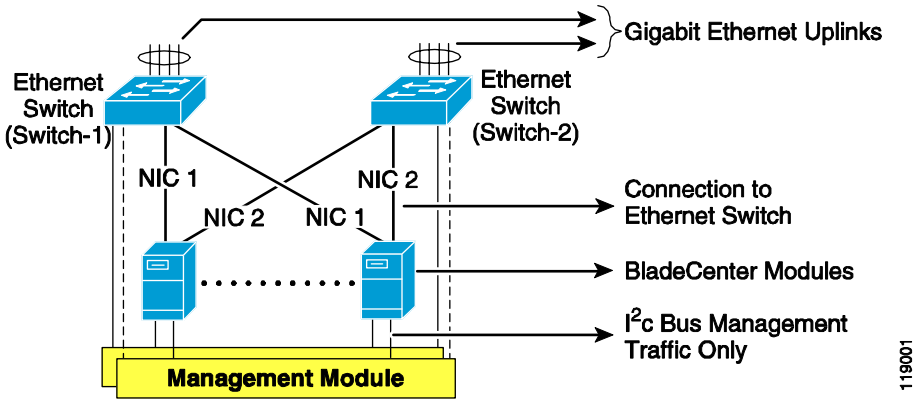
The Cisco IGESM integrates the Cisco industry-leading Ethernet switching technology into the IBM BladeCenter. For high availability and multi-homing, each IBM BladeCenter can be configured to concurrently support two pairs of Cisco IGESMs. The Cisco IGESM provides a broad range of Layer 2 switching features, while providing a seamless interface to SNMP-based management tools, such as CiscoWorks. The following switching features supported on the Cisco IGESM help provide this seamless integration into the data center network:

- Loop protection and rapid convergence with support for Per VLAN Spanning Tree (PVST+), 802.1w, 802.1s, BDPU Guard, Loop Guard, PortFast and UniDirectional Link Detection (UDLD)
- Advanced management protocols, including Cisco Discovery Protocol, VLAN Trunking Protocol (VTP), and Dynamic Trunking Protocol (DTP)

- Port Aggregation Protocol (PAgP) and Link Aggregation Control Protocol (LACP), for link load balancing and high availability
- Support for authentication services, including RADIUS and TACACS+
- Support for protection mechanisms, such as limiting the number of MAC addresses allowed, or shutting down the port in response to security violations

Each Cisco IGESM provides Gigabit Ethernet connectivity to each of the 14 blade slots in the BladeCenter and supplies four external Gigabit Ethernet uplink interfaces. You may install from one to four Cisco IGESMs in each BladeCenter. Figure 2-1 illustrates how the BladeCenter chassis provides Ethernet connectivity.

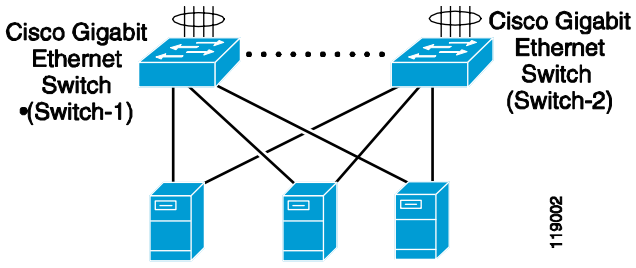
*Figure 2-1 BladeCenter Architecture for Ethernet Connectivity*




In Figure 2-1, two Ethernet switches within the BladeCenter chassis connect the blade server modules to external devices. Each Ethernet switch provides four Gigabit Ethernet links for connecting the BladeCenter to the external network. The uplink ports can be grouped to support the 802.3ad link aggregation protocol. In the illustrated example, each blade server is connected to the available Gigabit Ethernet network interface cards (NICs). NIC 1 on each blade server is connected to Cisco IGESM 1, while NIC 2 is connected to Cisco IGESM 2. The links connecting the blade server to the Cisco IGESM switches are provided by the BladeCenter chassis backplane.

Figure 2-2 provides a simplified logical view of the blade server architecture for data traffic. The dotted line between the two Cisco IGESMs shows the connectivity provided by the BladeCenter Management Module, which bridges traffic.

*Figure 2-2 Logical View of BladeCenter Chassis Architecture*



# Cisco IGESM Features

This section highlights information about protocols and features provided by Cisco IGESM that help integrate the BladeCenter into the Cisco Data Center Network Architecture and the IBM On-Demand Operating environment. This section includes the following topics:



## Spanning Tree

The Cisco IGESM supports various versions of the Spanning Tree Protocol (STP) and associated features, including the following:

- 802.1w
- 802.1s
- Rapid Per VLAN Spanning Tree Plus (RPVST+)
- Loop Guard
- Unidirectional Link Detection (UDLD)
- BPDU Guard

The 802.1w protocol is the standard for rapid spanning tree convergence, while 802.1s is the standard for multiple spanning tree instances. Support for these protocols is essential in a server farm environment for allowing rapid Layer 2 convergence after a failure in the primary path. The key benefits of 802.1w include the following:

- The spanning tree topology converges quickly after a switch or link failure.
- Convergence is accelerated by a handshake, known as the proposal agreement mechanism.
- There is no need to enable BackboneFast or UplinkFast.

In terms of convergence, STP algorithms based on 802.1w are much faster than traditional STP 802.1d algorithms. The proposal agreement mechanism allows the Cisco IGESM to decide new port roles by exchanging proposals with its neighbors.

With 802.1w, as with other versions of STP, bridge protocol data units (BPDUs) are still sent, by default, every 2 seconds (called the *hello time*). If three BPDUs are missed, STP recalculates the topology, which takes less than 1 second for 802.1w.

This seems to indicate that STP convergence time can be as long as 6 seconds. However, because the data center is made of point-to-point links, the only failures are physical failures of the networking devices or links. 802 1w is able to actively confirm that a port can safely transition to forwarding without relying on any timer configuration. This means that the actual convergence time is below *1 second* rather than 6 seconds.

The scenario where BPDUs are lost may be caused by unidirectional links, which can cause Layer 2 loops. To prevent this specific problem, you can use Loop Guard and UDLD. Loop Guard prevents a port from forwarding as a result of missed BPDUs, which might cause a Layer 2 loop that can bring down the network.

UDLD allows devices to monitor the physical configuration of fiber optic or copper Ethernet cables and to detect when a unidirectional link exists. When a unidirectional link is detected, UDLD shuts down the affected port and generates an alert. BPDU Guard prevents a port from being active in a spanning tree topology as a result of an attack or misconfiguration of a device connected to a switch port. The port that sees unexpected BPDUs is automatically disabled and must be manually enabled. This gives the network administrator full control over port and switch behavior.

The Cisco IGESM supports Per VLAN Spanning Tree (PVST) and a maximum of 64 spanning tree instances. RPVST+ is a combination of Cisco PVST Plus (PVST+) and Rapid Spanning Tree Protocol. Multiple Instance Spanning Tree (MST) adds Cisco enhancements to 802.1s. These protocols create a more predictable and resilient STP topology, while providing downward compatibility with simpler 802.s and 802.1w switches.

**Note** By default, the 802.1w protocol is enabled when running spanning tree in RPVST+ or MST mode.

## Traffic Monitoring

Cisco IGESM supports the following traffic monitoring features, which are useful for monitoring BladeCenter traffic in blade server environments:

- Switched Port Analyzer (SPAN)
- Remote SPAN (RSPAN)

SPAN mirrors traffic transmitted or received on source ports to another local switch port. This traffic can be analyzed by connecting a switch or RMON probe to the destination port of the mirrored traffic. Only traffic that enters or leaves source ports can be monitored using SPAN.

RSPAN enables remote monitoring of multiple switches across your network. The traffic for each RSPAN session is carried over a user-specified VLAN that is dedicated for that RSPAN session for all participating switches. The SPAN traffic from the source ports is copied onto the RSPAN VLAN through a reflector port. This traffic is then forwarded over trunk ports to any destination session that is monitoring the RSPAN VLAN.

## Link Aggregation Protocols

The Port Aggregation Protocol (PAgP) and Link Aggregation Control Protocol (LACP) help automatically create port channels by exchanging packets between Ethernet interfaces. PAgP is a Cisco-proprietary protocol that can be run only on Cisco switches or on switches manufactured by vendors that are licensed to support PAgP. LACP is a standard protocol that allows Cisco switches to manage Ethernet channels between any switches that conform to the 802.3ad protocol. Because the Cisco IGESM supports both protocols, you can use either 802.3ad or PAgP to form port channels between Cisco switches.

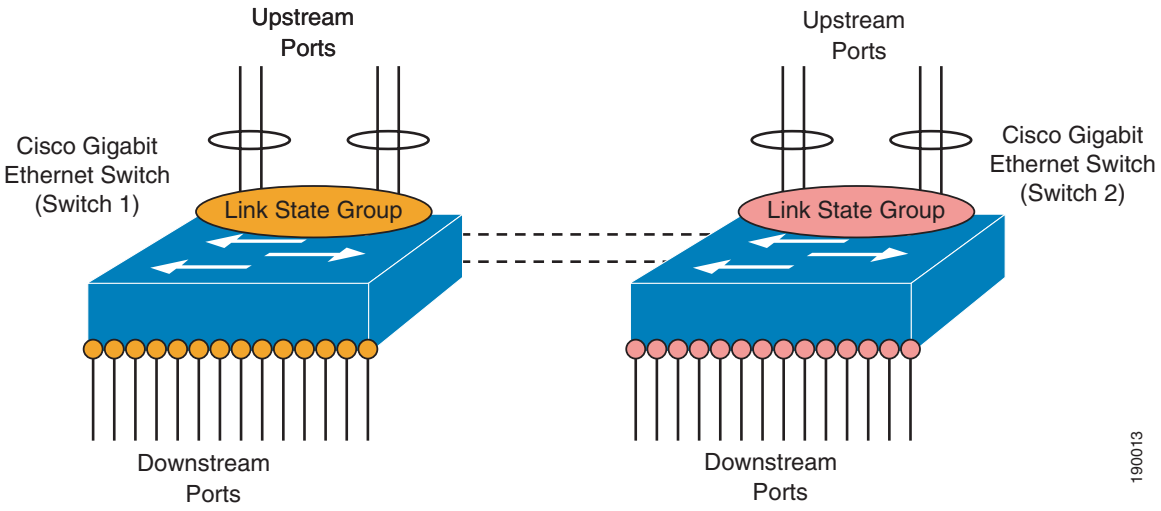
When using either of these protocols, a switch learns the identity of partners capable of supporting either PAgP or LACP and identifies the capabilities of each interface. The switch dynamically groups similarly configured interfaces into a single logical link, called a channel or aggregate port. The interface grouping is based on hardware, administrative, and port parameter attributes. For example, PAgP groups interfaces with the same speed, duplex mode, native VLAN, VLAN range, trunking status, and trunking type. After grouping the links into a port channel, PAgP adds the group to the spanning tree as a single switch port.

## Layer 2 Trunk Failover

Trunk failover is a high availability mechanism that allows the Cisco IGESM to track and bind the state of external interfaces with one or more internal interfaces. The four available Gigabit Ethernet uplink ports of the Cisco IGESM provide connectivity to the external network and can be characterized as "upstream" links. The trunk failover feature may track these upstream interfaces individually or as a port channel. Trunk failover logically binds upstream links together to form a link state group. The internal interfaces of the IGESM provide blade server connectivity and are referred to as "downstream" interfaces in the trunk failover configuration. This feature creates a relationship between the two interface types where the link state of the "upstream" interfaces defined in a link state group determines the link state of the associated "downstream" interfaces.

Figure 2-3 illustrates the logical view of trunk failover on the Cisco IGESM. The two external port channels of Switch-1 and Switch-2 are configured as upstream connections in a link state group local to the switch. The 14 internal blade server ports are downstream interfaces associated with each local group.

***Figure 2-3 Trunk Failover Logical View***




Trunk failover places downstream devices into the same link state, "up" or "down", based on the condition of the link state group. If an uplink or upstream failure occurs, the trunk failover feature places the downstream ports associated with those upstream interfaces into a link "down" or inactive state. When upstream interfaces are recovered, the related downstream devices are placed in an "up" or active state. An average failover and recovery time for network designs implementing the trunk failover feature is 3 seconds.

Consider the following when configuring the trunk failover on the Cisco IGESM:

- Internal ports (Gigabit Ethernet 0/1–14) may not be configured as "upstream" interfaces.
- External ports (Gigabit Ethernet 0/17–20) may not be configured as "downstream" interfaces.
- The internal management module ports (Gigabit Ethernet 0/15–16) may not be configured in a link state group.
- Trunk failover does not consider STP. The state of the upstream connections determines the status of the link state group not the STP state forwarding, blocking, and so on.
- Trunk failover of port channels requires that all of the individual ports of the channel fail before a trunk failover event is triggered.

- SPAN/RSPAN destination ports are automatically removed from the trunk failover link state groups.

# Using the IBM BladeCenter in the Data Center Architecture

The BladeCenter chassis provides a set of internal redundant Layer 2 switches for connectivity to the blade servers. Each blade server installed in the BladeCenter can use dual NICs connected to both Layer 2 switches. The BladeCenter can be also be deployed without redundant switches or dual-homed blade servers.

Figure 2-1 illustrates the physical connectivity of the BladeCenter switches and the Blade Servers within the BladeCenter, while the logical connectivity is shown in Figure 2-2. When using the Cisco IGESM, a BladeCenter provides four physical uplinks per Cisco IGESM to connect to upstream switches. Blade servers in the BladeCenter are dual-homed to a redundant pair of Cisco IGESMs.

BladeCenters can be integrated into the data center topology in various ways. The primary design goal is a fast converging, loop-free, predictable, and deterministic design, and this requires giving due consideration to how STP algorithms help achieve these goals.

This section describes the design goals when deploying blade servers and the functionality supported by the Cisco IGESM in data centers. It includes the following topics:

- High Availability
- Scalability
- Management

## High Availability

Traditionally, application availability has been the main consideration when designing a network for supporting data center server farms. Application availability is achieved through a highly available server and network infrastructure. For servers, a single point of failure is prevented through dual-homing. For the network infrastructure, this is achieved through dual access points, redundant components, and so forth.

When integrating the BladeCenter, the Cisco IGESM Layer 2 switches support unique features and functionality that help you achieve additional design considerations.

High availability, which is an integral part of data center design, requires redundant paths for the traffic to and from the server farm. In the case of a BladeCenter deployment, this means redundant blade server connectivity. The following are two areas on which to focus when designing a highly available network for integrating BladeCenters:

- High availability of the switching infrastructure provided by the Cisco IGESM
- High availability of the blade servers connected to the Cisco IGESM

### High Availability for the BladeCenter Switching Infrastructure

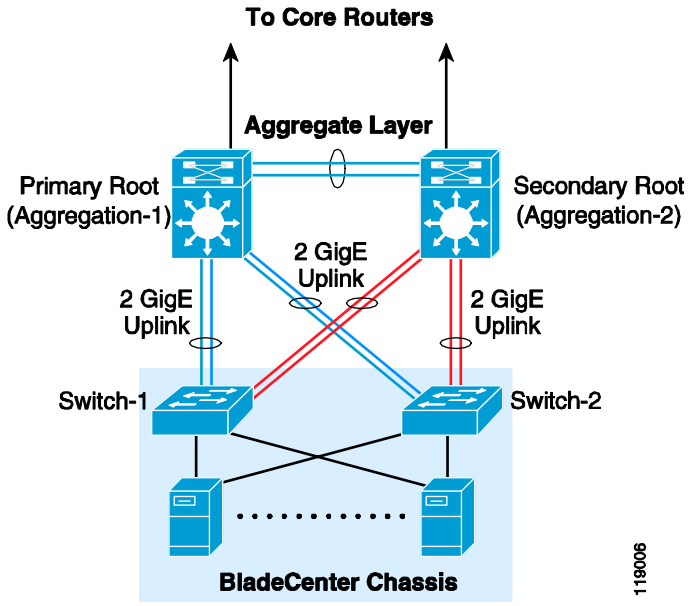
Redundant paths are recommended when deploying BladeCenters, and you should carefully consider the various failure scenarios that might affect the traffic paths. Each of the redundant BladeCenter Layer 2 switches provides a redundant set of uplinks, and the design must ensure fast convergence of the spanning tree topology when a failure in an active spanning tree link occurs. To this end, use the simplest possible topology with redundant uplinks and STP protocols that are compatible with the BladeCenter IGESMs and the upstream switches.

To create the redundant spanning tree topology, connect each of the BladeCenter IGESMs to a set of Layer 2/3 upstream switches that support RPVST+. To establish physical connectivity between the BladeCenter IGESMs and the upstream switches, dual-home each IGESM to two different upstream Layer 3 switches. This creates a deterministic topology that takes advantage of the fast convergence capabilities of RPVST+.

To ensure that the topology behaves predictably, you should understand its behavior in both normal and failure conditions. The recommended topology is described in more detail in Design and Implementation Details, page 2-13.

Figure 2-4 illustrates a fully redundant topology, in which the integrated Cisco IGESMs are dual-homed to each of the upstream aggregation layer switches. Each Cisco IGESM has a port channel containing two Gigabit Ethernet ports connected to each aggregation switch.

*Figure 2-4 Cisco IGESM Redundant Topology*




This provides a fully redundant topology, in which each BladeCenter switch has a primary and backup traffic path. Also notice that each Cisco IGESM switch has a deterministic topology in which RPVST+ provides a convergence time of less than one second after a failure. The environment is highly predictable because there is a single primary path used at all times, even when servers are dual-homed in active-standby scenarios.

**Note** The aggregation switches that provide connectivity to the BladeCenter are *multilayer* switches. Cisco does not recommend connecting a BladeCenter to Layer 2-only upstream switches.

## High Availability for the Blade Servers

Blade server high availability is achieved by multi-homing each blade to the integrated IGESMs employing the trunk failover feature. Multi-homing can consist of dual-homing each server to each of the Cisco IGESM switches, or using more than two interfaces per server, depending on the connectivity requirements.

Dual-homing leverages the NIC teaming features offered by the Broadcom chipset in the server NICs. These features support various teaming configurations for various operating systems. The following teaming mechanisms are supported by Broadcom:

- Smart Load Balancing
- Link Aggregation (802.3ad)
- Gigabit Cisco port channel

Smart Load Balancing is the only method of dual homing applicable to blade servers. The other two methods of teaming are not discussed in this document because they are not applicable. Although three teaming methods are supported, neither 802.3ad or Gigabit port channels can be used in the BladeCenter for high availability because the servers are connected to two different switches and the physical connectivity is dictated by the hardware architecture of the BladeCenter.

With Smart Load Balancing, both NICs use their own MAC addresses, but only the primary NIC MAC address responds to ARP requests. This implies that one NIC receives all inbound traffic. The outbound traffic is distributed across the two NICs based on source and destination IP addresses when the NICs are used in active-active mode.

The trunk failover feature available on the Cisco IGESM combined with the NIC teaming functionality of the Broadcom drivers provides additional accessibility to blade server resources. Trunk failover provides a form of "network awareness" to the NIC by binding the link state of upstream and downstream interfaces. The IGESM is capable of tracking the condition of its uplinks and placing associated "downstream" blade server ports in the same link state. If uplink failure occurs, the trunk failover feature disables the internal blade server ports, allowing a dual-homed NIC to converge using the high availability features of the NIC teaming driver. The trunk failover feature also recovers the blade server ports when uplink connectivity is re-established.

## Scalability

From a design perspective, Layer 2 adjacency also allows horizontal server farm growth. You can add servers to the same IP subnet or VLAN without depending on the physical switch to which they are connected, and you can add more VLANs/IP subnets to the server farm while still sharing the services provided by the aggregation switches.

Scaling the size of BladeCenters server farms depends on the following characteristics of the network:

- Physical port count at aggregation and access layers (the access layer being the Cisco IGESMs)
- Physical slot count of the aggregation layer switches

The following sections provide some guidance for determining the number of physical ports and physical slots available.

### Physical Port Count

Scalability, in terms of the number of servers, is typically determined by the number of free slots and the number of ports available per slot. With BladeCenter, this calculation changes because the blade servers are not directly connected to traditional external access layer or aggregation layer switches.

With BladeCenters, the maximum number of servers is limited by the number of BladeCenters and the number of ports in the upstream switches used to connect to the BladeCenters.

In the topology illustrated in Figure 2-1, for every 14 servers per BladeCenter, each aggregation switch needs to provide four Gigabit Ethernet ports (two to each Cisco IGESM).

The port count at the aggregation layer is determined by the number of slots multiplied by the number of ports on the line cards. The total number of slots available is reduced by each service module and supervisor installed.

Table 2-1 summarizes the total number of blade servers that can be supported for various line cards on a Cisco Catalyst 6500 switch on a per-line card basis. Keep in mind that the uplinks are staggered between two distinct aggregation switches, as shown in Figure 2-4.

*Table 2-1 BladeCenters Supported Based on Physical Port Count*

| Type of Line Card | Cisco IGESM per BladeCenter | Uplinks per Cisco IGESM | Total Uplinks | BladeCenters per Line Card |
|---|---|---|---|---|
| 8-port Gigabit Ethernet | 2 | 2 | 4 | 4 |
| | | 4 | 8 | 2 |
| | 4 | 2 | 8 | 2 |
| | | 4 | 16 | 1 |
| 16-port Gigabit Ethernet | 2 | 2 | 4 | 8 |
| | | 4 | 8 | 4 |
| | 4 | 2 | 8 | 4 |
| | | 4 | 16 | 2 |
| 48-port Gigabit Ethernet | 2 | 2 | 4 | 24 |
| | | 4 | 8 | 12 |
| | 4 | 2 | 8 | 12 |
| | | 4 | 16 | 6 |

## Slot Count

Your design should be flexible enough to quickly accommodate new service modules or BladeCenters without disruption to the existing operating environment. The slot count is an important factor in planning for this goal because the ratio of servers to uplinks dramatically changes as the number of BladeCenters increases.

This scaling factor is dramatically different than those found in traditional server farms where the servers are directly connected to access switches and provide very high server density per uplink. In a BladeCenter environment, a maximum of 14 servers is supported over as many as eight uplinks per BladeCenter. This creates the need for higher flexibility in slot/port density at the aggregation layer.

A flexible design must be able to accommodate growth in server farm services along with support for higher server density, whether traditional or blade servers. In the case of service modules and blade server scalability, a flexible design comes from being able to increase slot count rapidly without changes to the existing architecture. For instance, if firewall and content switching modules are required, the slot count on each aggregation layer switch is reduced by two.

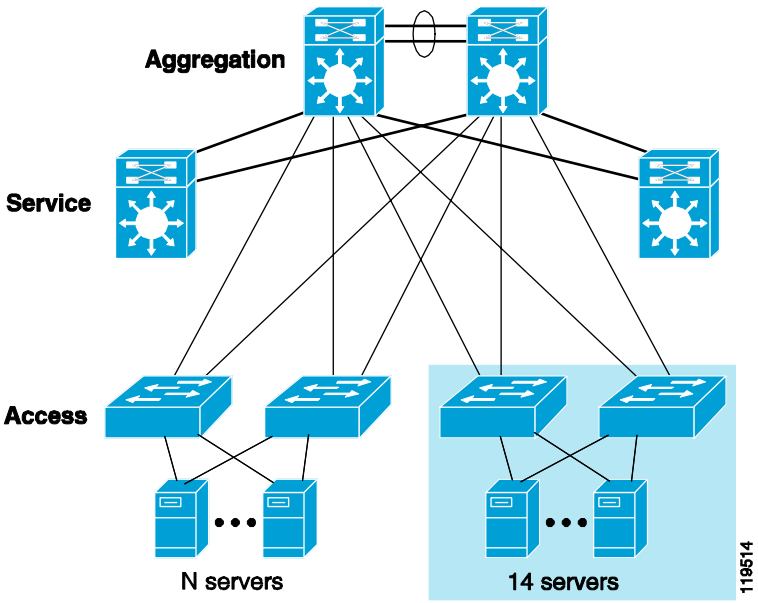
Cisco recommends that you start with a high-density slot aggregation layer and then consider the following two options to scale server farms:

- Use a pair of service switches at the aggregation layer.
- Use data center core layer switches to provide a scaling point for multiple aggregation layer switches.

Using service switches for housing service modules maintains the Layer 2 adjacency and allows the aggregation layer switches to be dedicated to provide server connectivity. This uses all available slots for line cards that link to access switches, whether these are external switches or integrated IGESMs. This type of deployment is illustrated in Figure 2-4.

Figure 2-5 illustrates traditional servers connected to access switches, which are in turn connected to the aggregation layer.

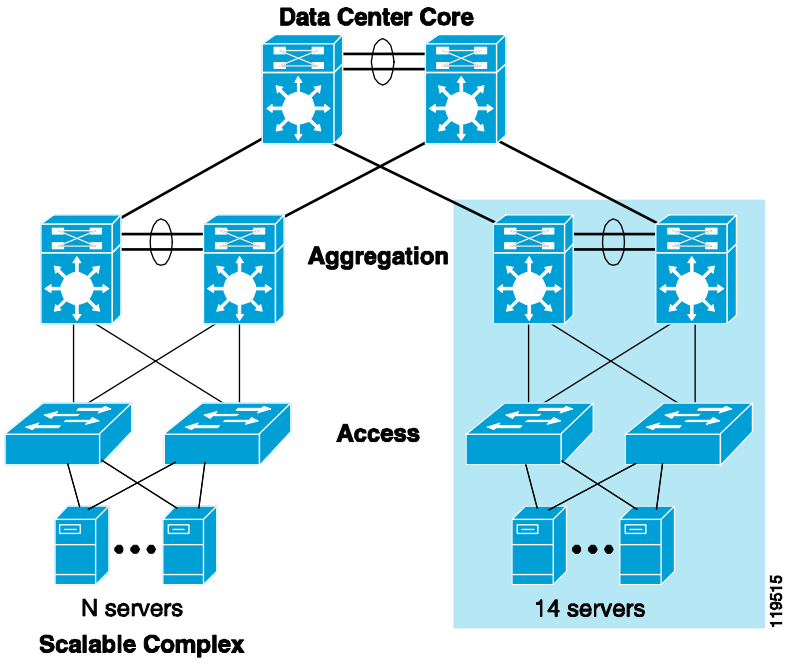
*Figure 2-5 Scaling With Service Switches*




Blade servers, on the other hand, are connected to the integrated IGESMs, which are also connected to the aggregation switches. The slot gained by moving service modules to the service layer switches lets you increase the density of ports used for uplink connectivity.

Using data center core layer switches allows scaling the server farm environment by sizing what can be considered a single module and replicating it as required, thereby connecting all the scalable modules to the data center core layer. Figure 2-6 illustrates this type of deployment.

*Figure 2-6 Scaling With Data Center Core Switches*




In the topology displayed in Figure 2-6, all service modules are housed in the aggregation layer switches. These service modules support the server farms that share the common aggregation switching, which makes the topology simple to implement and maintain. After you determine the scalability of a single complex, you can determine the number of complexes supported by considering the port and slot capacity of the data center core switches. Note that the core switches in this topology are Layer 3 switches.

## Management

You can use the BladeCenter Management Module to configure and manage the blade servers as well as the Cisco IGESMs within the BladeCenter without interfering with data traffic. To perform configuration tasks, you can use a browser and log into the management module.

Within the BladeCenter, the server management traffic (typically server console access) flows through a different bus, called the $I^2C$ bus. The $I^2C$ bus and the data traffic bus within the BladeCenter are kept separate.

The BladeCenter supports redundant management modules. When using redundant management modules, the backup module automatically inherits the configuration of the primary module. The backup management module operates in standby mode.

You can access the management module for configuring and managing the Cisco IGESMs using the following three methods, which are described in the following sections:

- Out-of-Band Management
- In-Band Management
- Serial/Console Port

## Out-of-Band Management

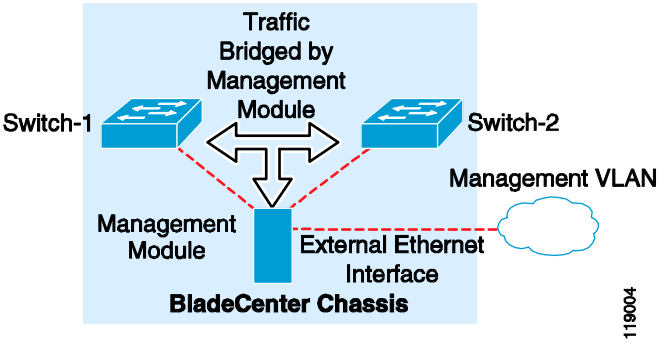
Out-of-band management refers to the common practice of dedicating a separate interface on each manageable device for management traffic.

Each management module in the BladeCenter supports an Ethernet interface, which is used to manage the blade servers and the Ethernet switches as well as the management module itself.

By default, the Ethernet switch management ports are placed in VLAN 1. It is very important that you put the management interface in a VLAN that is not shared by any of the blade server interfaces. The BladeCenter switch does not let you change the native VLAN on the interface connected to the management module.

Figure 2-7 presents a logical representation of an out-of-band management environment.

***Figure 2-7 Out-Of-Band Management***




Note that the dotted line shows control traffic between the external Ethernet interface and the Cisco IGESMs.

The management module Ethernet interface connects to the out-of-band management network. The management traffic destined for the Cisco IGESM comes through the management module Ethernet interface.

**Note** Do *not* configure VLAN 1 on any Cisco IGESM interface that is connected to a blade server. The Cisco IGESM management interface is in VLAN 1, by default, and all traffic from the blade server would be broadcast to VLAN 1.

## In-Band Management

With in-band management, Telnet or SNMP traffic uses the same path that is used by data, which limits the bandwidth available over uplinks. However, in-band management is common, especially where application management and network management traffic is separated in different VLANs or subnets.

Therefore, in addition to the VLAN 1 consideration, it is important to keep all management traffic on a different VLAN than non-control traffic.

## Serial/Console Port

Like other Cisco products, the Cisco IGESM has a serial port that can be used for management purposes. The serial port is typically connected to a terminal server through which the IGESMs can be managed remotely. You can establish a management session over a Telnet connection to the terminal server IP address and to the port to which the IGESM is connected. See the following URL for details about how

to set up console access to the switch through the serial/console port: http://www.cisco.com/en/US/docs/switches/lan/catalyst2950/software/release/12.1_9_ea1/configuration/guide/swadmin.html.

# Design and Implementation Details

This section provides design and implementation details for a server farm design that integrates BladeCenters. The following topics are included:

- Network Management Recommendations
- Layer 2 Looped Access Layer Design—Classic "V"
- Layer 2 Loop-Free Access Layer Design—Inverted "U"
- Configuration Details

## Network Management Recommendations

As described in the previous sections, there are various ways to manage the Cisco IGESM switches. Out-of-band management is the recommended option (see Figure 2-7) because it is simpler and management traffic can be kept on a separate VLAN. Additionally, setting the IP addresses on the IGESMs is a straightforward process using the graphic user interface (GUI) provided by the management module.

By default, the Cisco IGESM management ports are placed in VLAN 1. As discussed previously, it is very important that you put the management interface in a VLAN that is not shared by any of the blade server interfaces. The Cisco IGESM does not let you change the native VLAN on the management module interface. By using the default VLAN for the management module interfaces on each Cisco IGESM, management traffic and data traffic are kept in separate VLANs.

In-band management is not recommended for the following reasons:

- Management traffic must share the limited bandwidth between the aggregate switch and the Cisco IGESM switches.
- A loop is introduced by the management module.
- Broadcast traffic can conceivably overload the CPU.

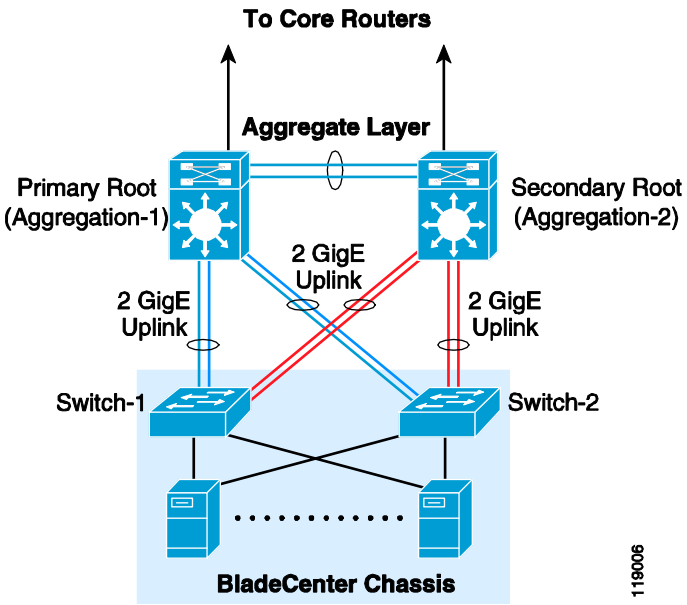
When you configure the BladeCenter switches for in-band management, the management interface is automatically assigned to the management VLAN (VLAN 1), which creates a loop. The management module has three interfaces and it bridges traffic received on these three interfaces. A loop is introduced by the management module because two of the interfaces are connected to the redundant Cisco IGESM switches. Although the Cisco IGESM has hardware filters to discard all the traffic between switches except spanning tree BPDUs and Cisco Discovery Protocol (CDP) packets, this mode is still not recommended. Although STP blocks potential loops, broadcast traffic can conceivably overload the CPU, which can lead to serious problems.

When you assign an IP address to the management interface of the Cisco IGESM, which for management purposes is a VLAN, the management module interface VLAN is automatically changed to the VLAN used along with IP address configuration. This helps ensure connectivity to the switches through the management module.

## Layer 2 Looped Access Layer Design—Classic "V"

Figure 2-8 illustrates the topology where the uplinks from the BladeCenter switches are distributed between two aggregate switches. This topology supports high availability by providing redundant Layer 2 links, so there is no single point of failure. The switches can be configured to provide a deterministic traffic path. In this topology, the BladeCenter switches must run STP to block Layer 2 loops.

*Figure 2-8 Uplinks Distributed Between Two Aggregate Switches*




The recommendation is to use RPVST+ because it provides convergence of less than one second in case of device or link failures. In addition to rapid convergence, RPVST+ incorporates many enhanced Cisco Layer 2 features, including BackboneFast and UplinkFast, which are used by default when you enable RPVST+.

To validate this topology, failure times were tested by sending traffic from the blade servers to a Layer 3 device on the aggregate switch at increasingly smaller intervals, and then measuring the number of packets lost. The following failure and recovery scenarios were tested:

- Uplink failure and recovery between Switch-1 and the primary root
- Uplink failure and recovery between Switch-2 and the primary root
- Switch-1 failure and recovery
- Switch-2 failure and recovery
- Primary root switch failure and recovery
- Secondary root switch failure and recovery

In most test cases, the failover and recovery times were a few hundred milliseconds. To allow a margin of error, the failover times can safely be rounded to one second. When the test case involves the failure of a switch that is the active HSRP device, the failover time is dependent on the HSRP failover time. Although HSRP can be configured to converge in sub-second times, a conservative estimate for recovery time when multiple components are involved is approximately five to six seconds.

These failover times are for Layer 2 and Layer 3 failures with HSRP at Layer 3. If the default gateway is on a different device, such as a firewall, the failover time for aggregate switch failure may change.

If a link fails within a port channel with two Ethernet links, the spanning tree topology does not change. The port channel simply stays up with a single link. This helps ensure that the BladeCenter traffic flow is not affected by the link failure.

The recommended topology provides redundant paths to BladeCenter traffic under all failure scenarios except for one case. This particular case is when all the links fail between a BladeCenter switch and the aggregation switches, and the NIC on the blade server is unaware of the uplink failure. The NIC teaming drivers cannot detect this condition and the servers are isolated until the links are restored. The trunk failover feature, available on the Cisco IGESM, addresses this situation. Trunk failover places blade server switch ports in a "down" state when their associated upstream uplinks fail. By doing so, the dual-homed server relies on the high availability features of the NIC teaming software to bypass the network failure and re-establish network connectivity in three seconds.

Cisco recommends alternating the active blade server interfaces on different Cisco IGESMs. This configuration helps prevent server isolation in the absence of trunk failover, and overall provides for better bandwidth utilization with this design. In addition, it places content switches in front of the server farm. The content switch can detect the failure or isolation of servers and can reroute requests to available resources.

It is also possible to monitor traffic on the Cisco IGESM switches in the topology shown in Figure 2-8. Under normal conditions, the backup links that are blocking can carry RSPAN traffic to the aggregate switches on a VLAN specifically used for mirrored traffic. This VLAN is configured only on Aggregation-2, on the Cisco IGESMs, and the backup link connecting these switches. This means that the topology is loop-free, and all ports are forwarding for this VLAN only. A network analysis probe should be attached to Aggregation-2. In this design, under failure conditions, the mirrored traffic shares the data traffic path.

**Note** The RSPAN VLAN is used only when it is required. For traffic mirroring, the Cisco IGESM switches require that one of the internal ports be configured as a reflector port, which means that one of the internal server ports is used for RSPAN. This is the recommended topology when a dedicated port for SPAN is not required.

To monitor traffic in this topology, perform the following steps:

**Step 1** Configure a VLAN for RSPAN and allow that VLAN on the backup port channel (blocking spanning tree link) on both the aggregate switch and the Cisco IGESM switch.

**Step 2** Configure traffic monitoring using the VLAN created in Step 1.

Given the current example, to configure traffic monitoring for the server connected to Gi0/5, enter the following commands:

```
monitor session 1 source interface Gi0/5
monitor session 1 destination remote vlan 300 reflector-port int Gi0/14
```

## Configuring the Aggregate Switches

Complete the following sequence of tasks on the aggregate switches:

1. VLAN configuration
2. RPVST+ configuration

3. Primary and secondary root configuration
4. Configuration of port channels between aggregate switches
5. Configuration of port channels between aggregate and Cisco IGESM switches
6. Trunking the port channels between aggregate switches
7. Configuration of default gateway for each VLAN

These tasks might be performed on a different device or a service module instead of the MSFC on the aggregate switch, depending on the architecture.

### Configuring the Cisco IGESM Switches

Complete the following sequence of tasks on the Cisco IGESM switches:

1. VLAN configuration
2. RPVST+ configuration
3. Configuration of port channels between the Cisco IGESM and aggregate switches
4. Trunking port channels between the Cisco IGESM and aggregate switches
5. Configuration of server ports on the Cisco IGESM
6. Configure trunk failover on the Cisco IGESM

Configuration Details, page 2-21 provides details for each of these steps. Most of these steps are required for either the primary recommended topology or the alternate topologies. In addition to these configuration steps, the following recommendations should be followed to ensure the successful integration of Cisco IGESM switches.

### Additional Aggregation Switch Configuration

**Step 1** Enable RootGuard on both the aggregate switch links that are connected to the Cisco IGESMs in the BladeCenter.

This prevents the Cisco IGESMs from assuming the STP root role by shutting down the interfaces in the aggregation switch that are connected to the Cisco IGESM. This safeguards against network meltdown in case of Cisco IGESM misconfiguration or misbehavior.

RootGuard can be enabled on both the aggregate switches using the **spanning-tree guard root** command in interface configuration mode. Enter this command on all the port channel interfaces between the aggregate switch and the Cisco IGESM switches.

**Step 2** Limit the VLANs on the port channel between the aggregate and Cisco IGESMs to those that are required.

Use the **switchport trunk allowed vlan** <*vlandID*> command on both the Cisco IGESM and aggregation switches in interface configuration mode. Enter this command on the Gigabit Ethernet interfaces and port channels.

### Additional Cisco IGESM Configuration

**Note** If BPDU Guard is enabled and a bridge protocol data unit (BPDU) is received on a port, the port is shut down. For this reason, do *not* enable BPDU Guard or BPDU filtering on the internal port connected to the management module because connectivity to the management module would be lost.

If connectivity to the management module is lost because BPDU Guard is enabled, you must reload the switch to recover from this condition. If the faulty configuration was saved, you must remove the connection from the management module external interface before reloading the switch.

**Step 1** Remove BPDU filtering.

By default, BPDU filter is enabled on all the internal ports of Cisco IGESM. To disable this, use the **spanning-tree bpdufilter disable** command in interface configuration mode.

**Step 2** Remove BPDU Guard.

To remove BPDU Guard, use the **spanning-tree bpduguard disable** command. However, BPDU guard can be enabled on the internal access ports connected to the blade servers.

**Step 3** Restrict VLANs on the port channel between the aggregate and Cisco IGESMs to those required.

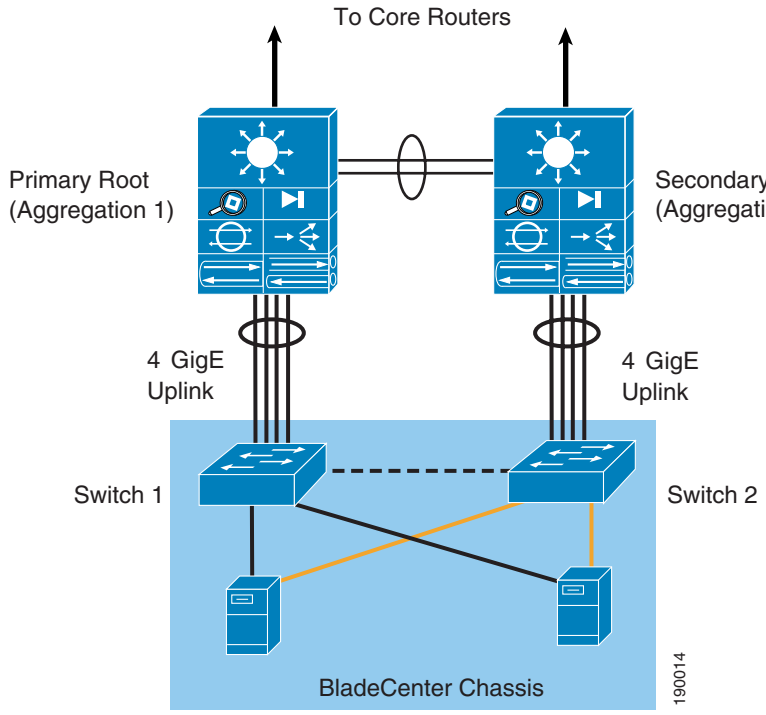
Use the **switchport trunk allowed vlan** <*vlanID*> command on both the Cisco IGESM and aggregation switches in interface configuration mode. This should be applied on the Gigabit Ethernet interfaces and port channels.

**Note** Do not use VLAN 1 on any of the interfaces on the Cisco IGESM connected to the blade servers. By default, the interface connected to the management module is in VLAN 1 and all the management traffic flows to the CPU of the Cisco IGESM.

## Layer 2 Loop-Free Access Layer Design—Inverted "U"

The topology in Figure 2-9 has the four individual uplinks from the BladeCenter switches connected to a single aggregate aggregation switch. This design is considered a loop-free topology. Switch-1 and Switch-2 have no interconnect; therefore, no loop exists. A redundant network path to the blade server NICs connected to each BladeCenter switch does not exist. Despite the absence of designed loops, Cisco recommends enabling spanning tree, preferably RPVST+, to manage potential loops created by misbehaving devices or human error.

*Figure 2-9 IGESM Uplinks Connected To One Aggregate Switch*




As noted with the previous design, a link failure between the BladeCenter switch and its associated aggregate switch isolates the blade server. The blade server is not aware of the network connectivity issues occurring beyond its own NIC. The trunk failover feature allows the Cisco IGESM and the blade server NIC teaming drivers to account for this scenario. The tests performed with this design achieved an average convergence of network traffic in three seconds.

The topology shown in Figure 2-9 has the advantage of higher bandwidth utilization when compared to other designs. Each Cisco IGESM uses the four Gigabit uplinks available. This design implies that blade server traffic that may pass through IGESM Switch-2 uses the aggregate inter-switch links to reach the primary root and the network services it may host. The links between the two aggregate switches should provide sufficient bandwidth to support all of the BladeCenters present in the data center. As a result, the configuration of the dual-homed blade server NICs becomes an important consideration.

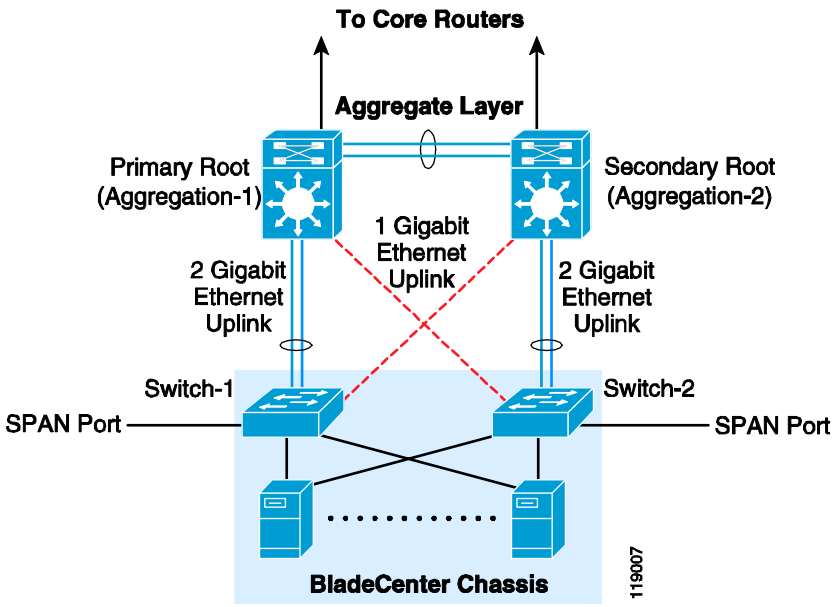
The primary interface assignment of a NIC team influences the traffic patterns within the data center. To avoid oversubscription of links between the aggregation switches and to simplify server deployments, assign the primary NIC on the IGESM homed to the primary root switch (Switch-1 in Figure 2-9). The oversubscription ratio of a fully-populated IBM BladeCenter with this NIC configuration is 3.5:1. The secondary NIC, assigned to the second IGESM (Switch-2 in Figure 2-10) is inactive unless activated by a failover. If oversubscribing the inter-switch links is not a concern, the blade server primary NIC interfaces may be evenly distributed between the two IGESMs, creating an oversubscription ratio on each switch of 1.75:1 that permits a higher bandwidth solution.

The configuration of this design requires the same steps as the recommended topology with the exception of redundant links to the aggregation layer from the IGESM.

Figure 2-10 and Figure 2-11 show two different topologies that provide a dedicated port for traffic monitoring. As mentioned in Network Management Recommendations, page 2-13, RSPAN requires a Cisco IGESM switch port to be configured as a reflector port. This port is then used for traffic monitoring instead of being available for a blade server. You can avoid using a valuable Cisco IGESM port by dedicating an external port to SPAN traffic.

**Note** This topology in which the link to the secondary root from Switch-2 is non-blocking is useful when performing any kind of maintenance function to the primary root that can affect the BladeCenter connectivity, thus providing the active server NICs on Switch-2 a primary forwarding path. This presumes that servers are both dual-homed and that half the servers are active on Switch-1 and the other half are active on Switch-2.

*Figure 2-10 First Topology with SPAN Ports*



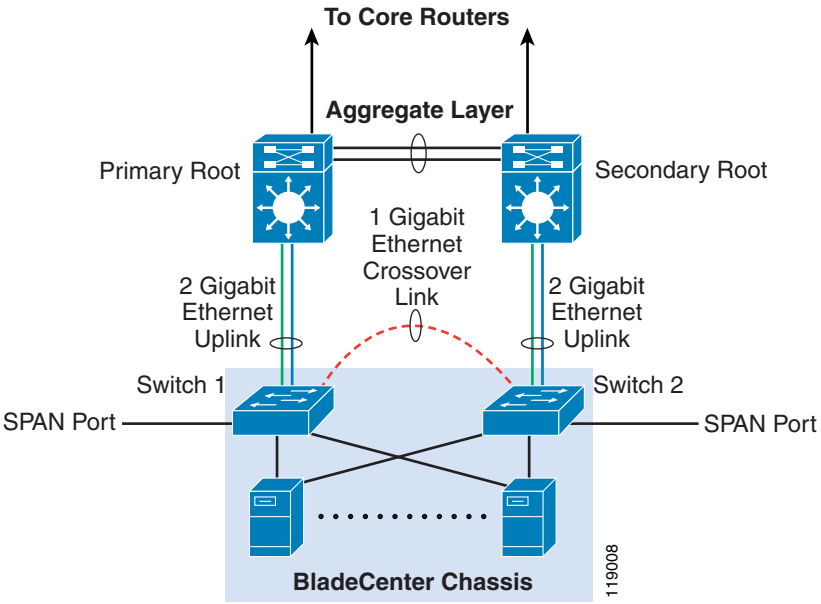

In this example, to monitor Blade Server 2 traffic, enter the following commands:

```
monitor session 1 source interface Gi0/2
monitor session 1 destination int Giga0/20
```

The topology presented in Figure 2-11 can be considered a variation of a topology in which the port channel is always non-blocking to the primary root, and the single link from either Cisco IGESM is to the secondary root, thus having a direct primary path from the primary root to each Cisco IGESM.

These topologies are recommended only if dedicated ports are required for monitoring traffic to and from the servers connected to Cisco IGESMs. In both topologies shown in Figure 2-10 and Figure 2-11, a dedicated port is used to monitor traffic on both the Cisco IGESM switches.

*Figure 2-11 Second Topology with SPAN Ports*




The disadvantage in these topologies is that in the event of device or link failures external to the BladeCenter, the redundant traffic path uses lower bandwidth links. You must monitor the link and device failures so you can quickly restore the higher bandwidth traffic path.

The difference between these two topologies is the traffic path during link or device failures. The advantage of the first topology (Figure 2-10) is that when the primary root switch fails, the traffic switches over to the secondary root directly (one hop versus two hops to get to the secondary root).

**Note** For the topology in Figure 2-10, you must change the path cost so that under normal circumstances, the traffic from Switch-2 in the BladeCenter always takes the port channel (2-port Gigabit Ethernet uplink) to the secondary root. See Configuration Details, page 2-21 for details about the commands required.

The second topology (Figure 2-11) saves some cabling effort and requires no modification to the spanning tree path cost. However, from the standpoint of providing an optimal traffic path, the first topology is recommended when dedicated ports are required for traffic monitoring.

For testing the failover times, the blade servers were dual-homed to both the switches and the RedHat Linux operating system was used. To test failover times for different links and devices, the following failure and recovery scenarios were tested:

- Uplink failure and recovery from Switch-1 to primary root
- Uplink failure and recovery from Switch-2 to secondary root
- Failure and recovery of Switch-1 and Switch-2 of BladeCenter
- Failure and recovery of the aggregation switches

In most cases, the failover time can still be rounded up to one second. However, as before, an active HSRP switch failure increases the failover time.

The best practices for this topology are the same as those described previously in Network Management Recommendations, page 2-13. For implementing this topology, the physical topologies may change and corresponding configuration changes will be required. However, the majority of the implementation details are similar to the recommended topology described previously.

The implementation steps for the recommended topology apply here as well. The differences in implementation are as follows:

- Inter-switch links can have single links, so port channel configuration is not required for every inter-switch link.
- The spanning tree path cost might have to be changed to ensure that the traffic follows the high bandwidth links under normal circumstances.

The details of the trunk configuration can be found in Configuration Details, page 2-21. For trunks between switches, ensure that only the required VLANs are allowed and that all traffic is tagged.

In the topology shown in Figure 2-10, you need to change the spanning tree path cost. Change the path cost of the trunk between Switch-2 and the primary root to a higher value (such as 8) on the Switch-2 interface. By changing this path cost, the traffic is forced to take the high bandwidth path if the traffic originates at Switch-2. Enter the following commands in global configuration mode to change the path cost.

```
interface interface-id
spanning-tree cost cost
spanning-tree vlan vlan-id cost cost
```

To verify the path cost, enter the following commands:

```
Show spanning-tree interface interface-id
Show spanning-tree vlan vlan-id
```

For the topology shown in Figure 2-11, you do not need to change the spanning tree path cost. In this topology, the only difference between that described in Figure 2-11 and in Network Management Recommendations, page 2-13, is that you must configure a trunk on the Gigabit Ethernet link between the Cisco IGESM switches and the aggregation switches.

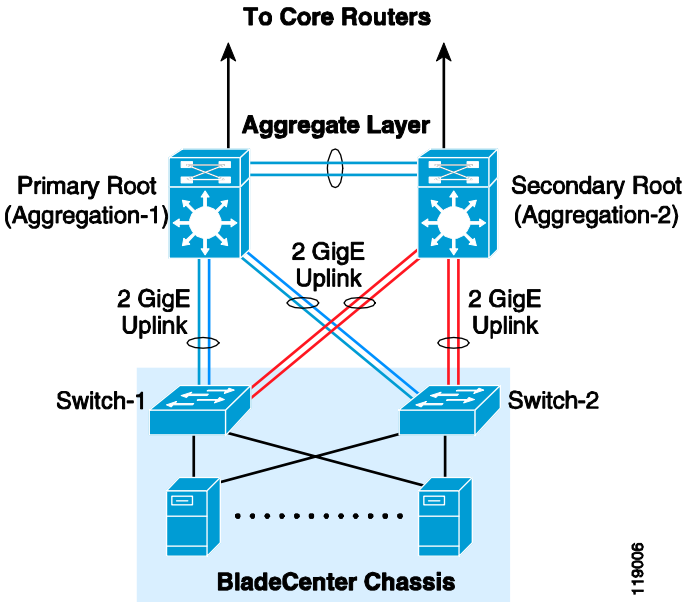
## Configuration Details

This section describes the configuration steps required for implementing the topologies discussed in this design guide. It includes the following topics:

- VLAN
- RPVST+
- Inter-Switch Link
- Port Channel
- Trunking
- Server Port
- Verifying Connectivity Between Cisco IGESMs
- Server Default Gateway
- Changing Spanning Tree Path Cost
- Layer 2 Trunk Failover

Figure 2-12 illustrates the topology to which this configuration applies and shows the primary root, secondary root, and where the port channel and link aggregation configurations are used.

*Figure 2-12 Example Topology*




## VLAN

Before configuring VLANs, you need to define the VTP mode for the switch. Enter the following commands to configure VTP:

```
vtp domain domain name
vtp mode transparent
```

Use the same VTP domain name everywhere in the data center. The configuration in this section covers only the server VLANs.

Other VLANs become involved when more services are added on top of the server VLANs. For more information about the configuration required for these additional services, see the *Data Center Infrastructure SRND* at the following URL: http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCI_SRND_2_5_book.html.

To create the VLANs, enter the following commands:

```
(config)# VLAN 10
(config-vlan)# name Blade1Vlan
```

VLANs 1–1002 are normal VLANs, while the VLANs numbered 1006 and higher are extended VLANs. After the VLANs are configured, name them appropriately and use the following command to place each VLAN in active state.

```
(config-vlan)# state active
```

## RPVST+

The following configuration details are for the aggregation switches. One aggregation switch is the primary root (by convention, the switch on the left) and the other aggregation switch is configured as the secondary root.

The examples in this design document show one aggregation switch (Aggregation-1) as the primary root and the other aggregation switch (Aggregation-2) as the secondary root for all the instances. This is because the design best practice is to not distribute traffic on uplinks to maintain a deterministic topology.

To configure RPVST+ in Cisco IOS Software Native Mode, enter the following command:

```
spanning tree mode rapid-pvst
```

A key recommendation is to have a single spanning tree topology for the VLANs in a set of redundant access and aggregation switches in the data center. If it is necessary to distribute the load through uplinks between access and aggregation switches, and to assign different priorities for the even and odd VLANs.

The next configuration step is to assign the root and the secondary root switches.

The configuration on Aggregation-1 for Cisco IOS Software Native Mode is as follows:

```
spanning-tree vlan 10,20,30 root primary
```

The configuration on Aggregation-2 for Cisco IOS Software Native Mode is as follows:

```
spanning-tree vlan 10,20,30 root secondary
```

With the **mac address reduction** option enabled, these commands assign the following election priorities:

- Root bridge priority—24576 (instead of 8192 without **mac address reduction** enabled)
- Secondary root bridge priority—28672 (instead of 16384 without **mac address reduction** enabled)
- Regular bridge priority—32768

**Note** With RPVST+, there is no need for UplinkFast and BackboneFast. Just configure RPVST+ in all the devices that belong to the same VTP domain.

### Inter-Switch Link

An inter-switch link here refers to the link between the switches and is not the same as ISL. There are two components in an inter-switch link:

- Port channel configuration
- Trunking configuration

The following inter-switch links are required for the topologies in this design guide:

- Inter-switch link between the aggregate switches (port-channel + trunk)
- Inter-switch link from both BladeCenter Cisco IGESMs to both aggregate switches (port-channel + trunk)
- In some topologies, you need to create a trunk link between the two Cisco IGESMs in the BladeCenter.

### Port Channel

To configure a port channel between the two aggregate switches, follow these guidelines:

- Use multiple ports from different line cards to minimize the risk of losing connectivity between the aggregation switches.
- Configure LACP active on Aggregation-1.

- Configure LACP passive on Aggregation-2.

The following example shows the channel configuration between the ports Giga1/1, Giga1/2, Giga6/1, and Giga6/2 for Aggregation-1.

```
interface GigabitEthernet1/1
description to_Aggregation-2
channel-group 2 mode active
channel-protocol lacp
interface GigabitEthernet1/2
description to_Aggregation-2
channel-group 2 mode active
channel-protocol lacp
interface GigabitEthernet6/1
description to_Aggregation-2
channel-group 2 mode active
channel-protocol lacp
interface GigabitEthernet6/2
description to_Aggregation-2
channel-group 2 mode active
channel-protocol lacp
```

The configuration for Aggregation-2 is the same with the exception that the channel mode is passive, configured with the following command*:*

```
channel-group 2 mode passive
```

For more information about configuring port channels, see the following URL: http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.1E/native/configuration/guide/channel.html.

## Trunking

To configure trunking between the two switches, follow these guidelines:

- Do not configure the trunks to allow all VLANs. Allow only those that are used.
- Use 802.1q trunking.
- Tag all VLANs over a trunk from the aggregation switches.

To define the VLANs allowed on a trunk, enter the following command:

```
switchport trunk allowed vlan 10,20
```

To modify the list of the VLANs allowed on a trunk, enter the following commands in Cisco IOS Software Native Mode:

```
switchport trunk allowed vlan add vlan number
switchport trunk allowed vlan remove vlan number
```

The recommended trunk encapsulation is 802.1q because it is the standard. The configuration in Catalyst 6500 IOS is as follows:

```
switchport trunk encapsulation dot1q
```

With some software versions, 802.1q is the default encapsulation and the **dot1q** option is not required.

You can force a port to be a trunk by entering the following command:

```
switchport mode trunk
```

This mode puts the port into permanent trunk mode and sends DTP frames to turn the neighboring port into a trunk as well. If the trunk does not form, verify the VTP domain configuration. VTP domain names must match between the neighboring switches.

The port channels that connect the aggregation switch to the BladeCenter switches are also trunked and should have Root Guard configured as shown in the following configuration. This configuration also shows a port channel trunking from Aggregation-1 to Switch-1

```
interface GigabitEthernet6/5
 description to-BladeCenterSW1/1
 no ip address
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport trunk allowed vlan 10,20
 switchport mode trunk
 spanning-tree guard root
 channel-group 5 mode active
 channel-protocol lacp

interface GigabitEthernet6/6
 description to-BladeCenterSW1/1
 no ip address
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport trunk allowed vlan 10,20
 switchport mode trunk
 spanning-tree guard root
 channel-group 5 mode active
 channel-protocol lacp
```

**Note** Configuring Root Guard on the port channel interface between the two Catalyst 6500 switches ensures that the primary root (Aggregation-1) is always the root for all the VLANs as long as the switch is up and running.

## Server Port

When a blade server is inserted into the BladeCenter, the blade server NIC attaches to specific Gigabit Ethernet interfaces on both BladeCenter switches based on the slot into which the blade is inserted. For example, when a blade server is inserted into slot 8, it is connected to Gigabit Ethernet interface 0/8 on both switches.

Trunking and PortFast are enabled by default on the access ports. It is also useful to trunk the server ports if trunking is configured on the blade server NICs. If trunking is not enabled on the server NICs (whether teamed or not), you can change the configuration to non-trunking and configure the port in access mode.

In addition, BPDU filtering is enabled by default. If the server drivers and operating systems do not bridge BPDUs and do not have to see BPDUs, there is no need to change this default configuration. However, if you enable BPDU Guard and disable BPDU filtering, the BPDUs are allowed to pass through from the Cisco IGESM interface to the blade server modules, and if the blade server modules bridge the BPDUs, BPDU Guard shuts down the port.

**Note** Do not enable BPDU Guard or BPDU filtering on the internal port connected to the management module.

If BPDU Guard is enabled and a BPDU is received on this interface, the port shuts down and connectivity to the management module is lost. To recover from this condition, you have to reload the switch. If the faulty configuration was saved, you must remove the connection from the management module external interface before reloading the switch.

Enable trunk failover on the "downstream" internal blade server ports and assign the port to a specific trunk failover group. Port security can also be enabled on the access ports. The new non-trunking (access port) configuration is as follows:

```
interface GigabitEthernet0/8
 description blade8
 switchport access vlan 20
 switchport mode access
 switchport port-security maximum 3
 switchport port-security aging time 20
 link state group 1 downstream
 spanning-tree portfast
 spanning-tree bpduguard enable
 no cdp enable
end
```

**Note** No explicit port speed settings or duplex settings are shown here because auto-negotiation is reliable between the blade servers and the BladeCenter Cisco IGESM. However, if the environment requires ports to come up faster, configure explicit port speed and duplex settings.

For more information about PortFast, BPDU Guard, and TrunkFast, see the following URL: http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.1E/native/configuration/guide/stp_enha.html.

By default, the Cisco IGESM port connected to the management module is in VLAN 1. Putting any of the blade servers in this VLAN causes broadcast traffic to show up in the Cisco IGESM CPU queue. This might overload the CPU and cause other undesired results. Also, as mentioned previously, the management module bridges traffic coming from its external port to both the Cisco IGESM switches.

## Verifying Connectivity Between Cisco IGESMs

This section explains how to verify the connectivity between two Cisco IGESM switches in the same chassis.

To confirm the connectivity between the two switches, enter the **show cdp neighbor** command on either of the switches. The following shows sample output from this command:

```
Switch-1# show cdp neighbors Gigabit Ethernet 0/15 detail
-------------------------
Device ID: Switch-1.example.com
Entry address(es):
  IP address: 192.26.208.14
Platform: cisco OS-Cisco IGESM-18,  Capabilities: Switch IGMP
Interface: GigabitEthernet0/15,  Port ID (outgoing port): GigabitEthernet0/15
Holdtime : 134 sec

Version :
Cisco Internetwork Operating System Software
IOS (tm) Cisco IGESM Software (Cisco IGESM-I6Q4L2-M), Version 12.1(0.0.42)AY, CISCO DEVELO
PMENT TEST VERSION
Copyright (c) 1986-2004 by Cisco Systems, Inc.
Compiled Mon 08-Mar-04 10:20 by antonino
advertisement version: 2
Protocol Hello:  OUI=0x00000C, Protocol ID=0x0112; payload len=27, value=0000000
0FFFFFFFF010221FF000000000000000ED7EE0900FF0000
VTP Management Domain: 'blade-centers'
Native VLAN: 1
Duplex: full
```

**Note** Do not use the management VLAN for any blade server module. Always keep the management VLAN and the blade server VLANs separate.

## Server Default Gateway

The default gateway for the blade servers is typically configured on a Layer 3 device, which can be a Firewall Service Module (FWSM) or the Multilayer Switch Feature Card (MSFC).

Assuming that the default gateway is on the MSFC, you can use HSRP to configure a highly available Layer 3 interface. The Layer 3 interface is on the aggregation switch. A Layer 3 interface exists on the aggregation switch for each server VLAN on the BladeCenter Cisco IGESM.

You also configure the server VLAN on the aggregate switch, and the trunks between the two aggregate switches carry this server VLAN. The VLANs on these aggregate switch trunks carry HSRP heartbeats between the active and standby HSRP interfaces. The HSRP configuration on aggregate1 (HSRP Active) is shown below.

```
interface Vlan10
 description BladeServerFarm1
 ip address 10.10.10.2 255.255.255.0
 no ip redirects
 no ip proxy-arp
 arp timeout 200
 standby 1 ip 10.10.10.1
 standby 1 timers 1 3
 standby 1 priority 110
 standby 1 preempt delay minimum 60
 standby 1 authentication cisco
end

interface Vlan20
 description BladeServerFarm2
 ip address 10.10.20.2 255.255.255.0
 no ip redirects
 no ip proxy-arp
 arp timeout 200
 standby 1 ip 10.10.20.1
 standby 1 timers 1 3
 standby 1 priority 110
 standby 1 preempt delay minimum 60
 standby 1 authentication cisco
end

The configuration on Aggregation-2 (HSRP Standby) is as shown below.
interface Vlan10
 description BladeServerFarm1
 ip address 10.10.10.3 255.255.255.0
 no ip redirects
 no ip proxy-arp
 arp timeout 200
 standby 1 ip 10.10.10.1
 standby 1 timers 1 3
 standby 1 priority 100
 standby 1 preempt delay minimum 60
 standby 1 authentication cisco
end

interface Vlan20
 description BladeServerFarm2
 ip address 10.10.20.3 255.255.255.0
```

```
 no ip redirects
 no ip proxy-arp
 arp timeout 200
 standby 1 ip 10.10.20.1
 standby 1 timers 1 3
 standby 1 priority 100
 standby 1 preempt delay minimum 60
 standby 1 authentication cisco
end
```

### Changing Spanning Tree Path Cost

This section is applicable to the topology introduced in Figure 2-10 only.

Changing spanning tree parameters is not required in most cases. However, when dedicated ports are used for traffic monitoring, change the path cost to force the traffic through higher bandwidth links.

For example, to implement the topology shown in Figure 2-10, on the Switch-2 interface connected to the primary root, change the path cost to 8. Use the following commands in global configuration mode to change the path cost.

```
interface interface-id
spanning-tree cost cost
spanning-tree vlan vlan-id cost cost
```

To verify the path cost, enter the following commands:

```
Show spanning-tree interface interface-id
Show spanning-tree vlan vlan-id
```

### Layer 2 Trunk Failover

The trunk failover feature may track an upstream port or a channel. To assign an interface to a specific link state group, use the following command in the interface configuration sub mode:

```
link state group <1-2> upstream
```

**Note** Gigabit Ethernet interfaces 0/17–20 may be configured only as “upstream” devices.

Enable the Trunk Failover feature for the downstream ports of the internal blade server interfaces for a specific link state group.

```
interface GigabitEthernet0/1
 description blade1
 link state group <1-2> downstream

interface GigabitEthernet0/2
 description blade2
 link state group <1-2> downstream
```

**Note** Gigabit Ethernet interfaces 0/1–14 may be configured only as “downstream” devices.

Globally enable the trunk failover feature for a specific link state group:

```
link state track <1-2>
```

To validate the trunk failover configuration, use the following command:

```
show link state group detai
```

# Cisco Gigabit Ethernet Switch Module for the HP BladeSystem

This section provides best design practices for deploying the Cisco Gigabit Ethernet Switch Modules (CGESM) for the HP BladeSystem p-Class enclosures within the Cisco Data Center Networking Architecture. This document describes the internals of the blade enclosure and CGESM and explores various methods of deployment. It includes the following sections:



## Cisco Gigabit Ethernet Switching Module

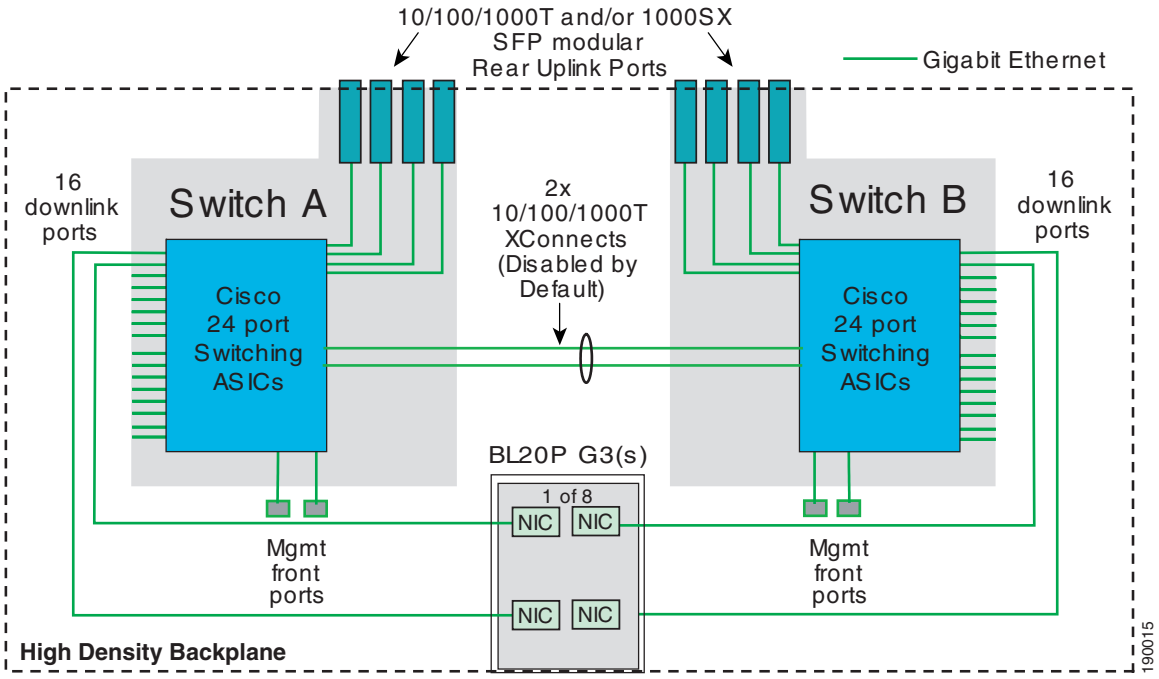
This section briefly describes the Cisco Gigabit Ethernet Switching Module (CGESM) and explains how the blade servers within the HP BladeSystem are physically connected to the switching module.

The CGESM provides enhanced Layer 2 services (known as L2+ or Intelligent Ethernet) to the HP BladeSystem p-Class. The CGESM extends the capabilities of a Layer 2 Ethernet switch to include Cisco proprietary protocols, ACLs, and QoS based on Layer 3 information. With SNMP, CLI, or HTTP management options available and a robust set of IOS switching features, the CGESM naturally integrates into the data center environment. The following features highlight this capacity:

- Loop protection and rapid convergence with support for Trunk Failover, Per VLAN Spanning Tree (PVST+), 802.1w, 802.1s, BDPU Guard, Loop Guard, PortFast, UplinkFast, and UniDirectional Link Detection (UDLD)
- Advanced management protocols, including Cisco Discovery Protocol (CDP), VLAN Trunking Protocol (VTP), and Dynamic Trunking Protocol (DTP)
- Port Aggregation Protocol (PAgP) and Link Aggregation Control Protocol (LACP) for link load balancing and high availability
- Support for authentication services, including RADIUS and TACACS+ client support
- Support for protection mechanisms, such as limiting the number of MAC addresses allowed, or shutting down the port in response to security violations

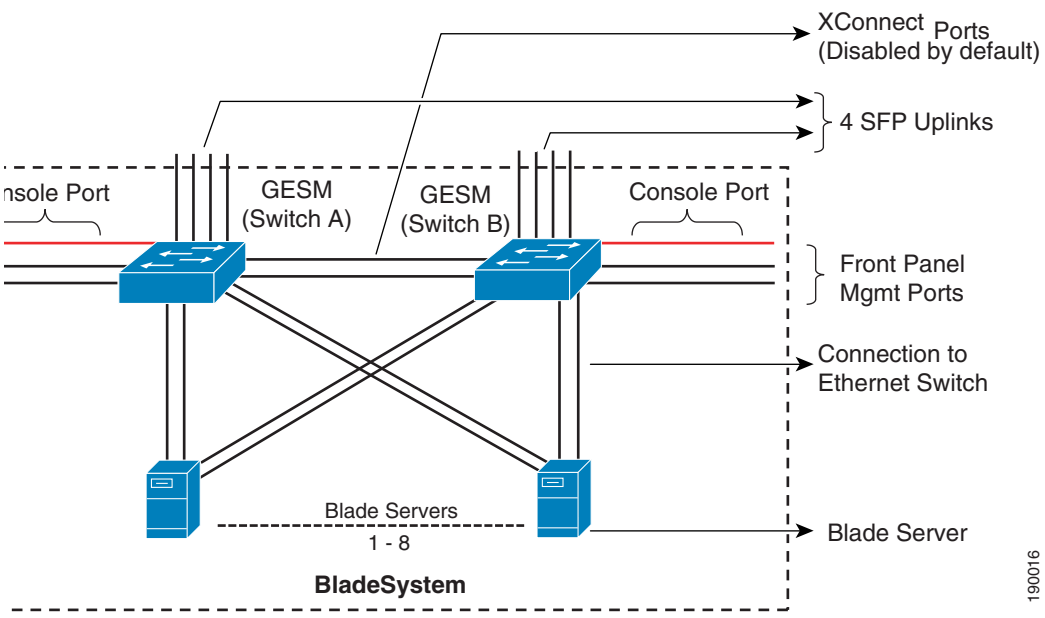
The HP BladeSystem p-Class enclosure consists of eight server bays and two network-interconnect bays. Figure 2-13 shows the BladeSystem p-Class architecture using two CGESMs housed in the network interconnect bays and eight BL20P G3 servers.

*Figure 2-13 BladeSystem p-Class Switch Architecture*




The HP BladeSystem p-Class backplane provides power and network connectivity to the blades. The interconnect bays house a pair of CGESMs, which provide a highly available and multi-homed environment where each server blade is Gigabit attached to each CGESM. Figure 2-14 illustrates how the HP BladeSystem p-Class logically provides Ethernet connectivity.

*Figure 2-14 Blade Enclosure Ethernet Connectivity*



**Note** Figure 2-14 is based on the use of the HP BladeSystem p-Class using BL20P G3 servers. The remainder of this document uses the BL20P G3 server for all figures.

In Figure 2-14, two CGESMs within the blade enclosure connect the blade server modules to external network devices such as aggregation layer switches. Each Ethernet switch provides six external Ethernet ports for connecting the blade enclosure to the external network. Four SFP ports provide 1000 Base-SX and 10/100/1000 Base-T links on the rear of the enclosure and two 10/100/1000 Base-T ports provide connectivity on the front panel. All six of these ports can be grouped to support the 802.3ad link aggregation protocol. In Figure 2-14 above, each blade server is connected to the backplane via the available Gigabit Ethernet network interface cards (NICs). The number of NICs on each blade server varies. Table 2-2 provides more detail on the connectivity options available with each HP blade server and the maximum number of blade servers a single enclosure can support.

**Note** This list is not comprehensive; more detail on HP blade server connectivity can be found at the following URL: http://www.hp.com.

*Table 2-2 Blade Server Options*

| Blade Server | Maximum Number of Server Blades per Enclosure | NICs Available |
|---|---|---|
| BL20P G2 | 8 | 3 10/100/1000T NICs<br>1 dedicated iLO interface |
| BL20P G3 | 8 | 4 10/100/1000T NICs<br>1 dedicated iLO interface |
| BL30P | 16 | 2 10/100/1000T NICs<br>1 dedicated iLO interface |
| BL40P | 2 | 5 x 10/100/1000T NICs<br>2 Slots for SAN connectivity<br>1 dedicated iLO interface |

**Note** In Table 2-2, “iLO” refers to the Integrated Lights-Out interface. It supports the iLO management subsystem that resides on each server blade. For more information on the iLO system, see Management, page 2-42.

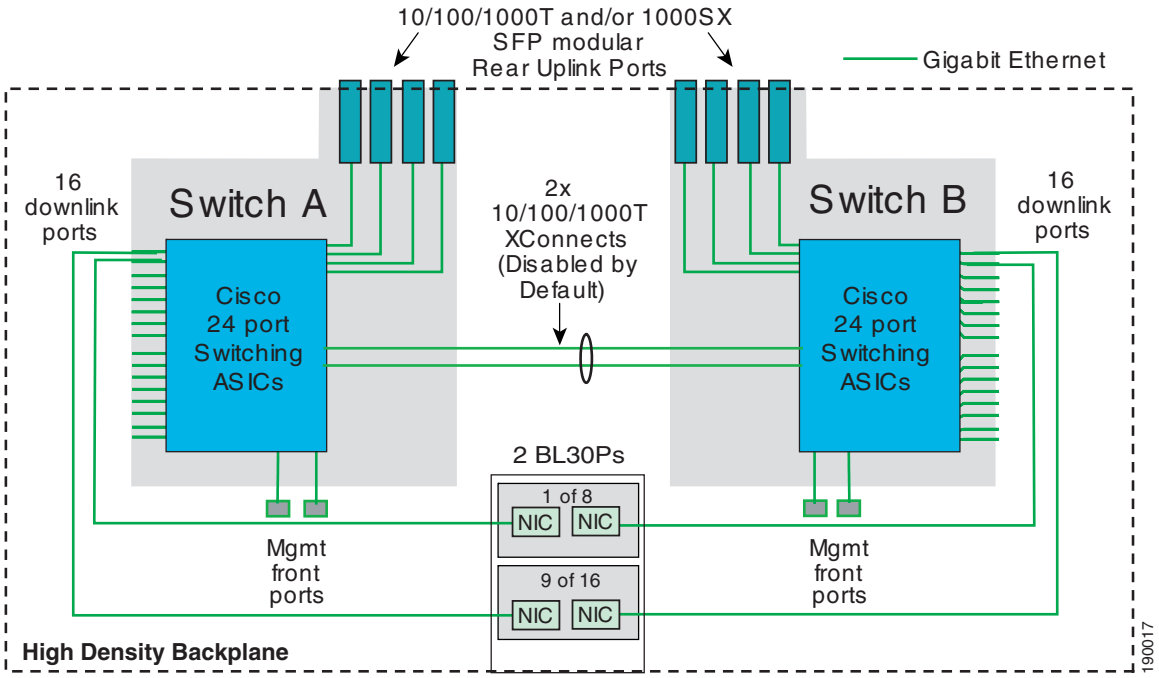
In Figure 2-13 and Figure 2-14, two NICs on each blade server connect to CGESM A and CGESM B. The blade servers connect to the CGESM switches over the HP BladeSystem p-Class backplane. There are sixteen 10/100/1000 internal ports on each CGESM dedicated to the blade server modules.

Figure 2-13 and Figure 2-14 also show two internal 10/100/1000 ports interconnecting the two blade enclosure switches over the backplane. These ports are disabled by default, but are configurable to carry all traffic types between the two blade enclosure switches. These ports support trunking and can be configured as channels.

**Note** In Figure 2-13, if the BL20p G1 and G2 servers are used, each server dedicates one NIC connected to the CGESM B side to the iLO port. This NIC is capable only of 100 MB. The enclosure with an enhanced backplane provides blade server connectivity to an embedded iLO module on the backplane. See Management, page 2-42 for more details.

The HP BladeSystem p-Class enclosure with enhanced backplane consists of eight server bays and two network-interconnect bays. The HP BladeSystem p-Class sleeve option allows the enclosure to support 16 Proliant BL30p servers or two BL30p servers per bay. Figure 2-15 illustrates the Gigabit Ethernet connectivity of an enhanced backplane enclosure and 16 BL30p servers.

***Figure 2-15 HP BladeSystem p-Class with 16 Servers***




For more information about the HP BladeSystem p-Class, see the following URL: http://h18004.www1.hp.com/products/servers/proliant-bl/p-class/documentation.html.

# CGESM Features

This section provides information about the protocols and features provided by the CGESM that help to integrate the HP BladeSystem p-Class enclosure into the Cisco Data Center Network Architecture. This section includes the following topics:

- Spanning Tree
- Traffic Monitoring
- Link Aggregation Protocols
- Layer 2 Trunk Failover

## Spanning Tree

The CGESM supports various versions of the Spanning Tree Protocol (STP) and associated features, including the following:

- Rapid Spanning Tree (RSTP) based on 802.1w
- Multiple Spanning Tree (MST) based on 802.1s with 802.1w
- Per VLAN Spanning Tree Plus (PVST+)
- Rapid Per VLAN Spanning Tree Plus (RPVST+)
- Loop Guard
- Unidirectional Link Detection (UDLD)
- BPDU Guard
- PortFast
- UplinkFast (Cisco proprietary enhancement for 802.1d deployments)
- BackboneFast (Cisco proprietary enhancement for 802.1d deployments)

The 802.1w protocol is the standard for rapid spanning tree convergence, while 802.1s is the standard for multiple spanning tree instances. Support for these protocols is essential in a server farm environment for allowing rapid Layer 2 convergence after a failure occurs in the primary path. The key benefits of 802.1w include the following:

- The spanning tree topology converges quickly after a switch or link failure.
- Convergence is accelerated by a handshake, known as the proposal agreement mechanism.

**Note** There is no need to enable BackboneFast or UplinkFast

In terms of convergence, STP algorithms based on 802.1w are much faster than the traditional STP 802.1d algorithms. The proposal agreement mechanism allows the CGESM to decide new port roles by exchanging proposals with its neighbors.

With 802.1w, as with other versions of the STP, BPDUs are by default sent every two seconds (called the *hello time*). If three BPDUs are missed, STP recalculates the topology, which takes less than one second for 802.1w.

This seems to indicate that STP convergence time can be as long as six seconds. However, because the data center is made of point-to-point links, the only failures are physical failures of the networking devices or links. 802.1w is able to actively confirm that a port can safely transition to forwarding without relying on any timer configuration. This means that the actual convergence time is below *one second* rather than six seconds.

A scenario where BPDUs are lost may be caused by unidirectional links, which can cause Layer 2 loops. To prevent this problem, you can use Loop Guard and UDLD. Loop Guard prevents a port from forwarding as a result of missed BPDUs, which might cause a Layer 2 loop that can bring down the network.

UDLD allows devices to monitor the physical configuration of fiber optic or copper Ethernet cables and to detect when a unidirectional link exists. When a unidirectional link is detected, UDLD shuts down the affected port and generates an alert. BPDU Guard prevents a port from being active in a spanning tree topology as a result of an attack or misconfiguration of a device connected to a switch port. The port that sees unexpected BPDUs is automatically disabled and must then be manually enabled. This gives the network administrator full control over port and switch behavior.

The CGESM supports PVST and a maximum of 128 spanning tree instances. RPVST+ is a combination of Cisco PVST Plus (PVST+) and Rapid Spanning Tree Protocol. RPVST+ provides the flexibility of one spanning tree instance per VLAN and fast convergence benefits of 802.1w. Multiple Spanning Tree (MST) allows the switch to map several VLANs to one spanning tree instance, reducing the total number of spanning tree topologies the switch processor must manage. A maximum of 16 MST instances are supported. In addition, MST uses 802.1w for rapid convergence. MST and RPVST+ create a more predictable and resilient spanning tree topology, while providing downward compatibility for integration with devices that use 802.1d and PVST+ protocols.

**Note** The 802.1w protocol is enabled by default when running spanning tree in RPVST+ or MST mode on the CGESM. CGESM enables PVST+ for VLAN 1 by default.

Spanning tree uses the path cost value to determine the shortest distance to the root bridge. The port path cost value represents the media speed of the link and is configurable on a per interface basis, including EtherChannels. To allow for more granular STP calculations, enable the use of a 32-bit value instead of the default 16-bit value. The *longer* path cost better reflects changes in the speed of channels and allows STP to optimize the network in the presence of loops.

**Note** The CGESM supports IEEE 802.1t, which allows for spanning tree calculations based on a 32-bit path cost value instead of the default 16 bits. For more information about the standards supported by the CGESM, see *Cisco Gigabit Ethernet Switch Module (CGESM) Overview.*

## Traffic Monitoring

The CGESM supports the following traffic monitoring features, which are useful for monitoring blade enclosure traffic in data center environments:

- Switched Port Analyzer (SPAN)
- Remote SPAN (RSPAN)

SPAN mirrors traffic transmitted or received on source ports or source VLANs to another local switch port. This traffic can be analyzed by connecting a switch or RMON probe to the destination port of the mirrored traffic. Only traffic that enters or leaves source ports or source VLANs can be monitored using SPAN.

RSPAN enables remote monitoring of multiple switches across your network. The traffic for each RSPAN session is carried over a user-specified VLAN that is dedicated for that RSPAN session for all participating switches. The SPAN traffic from the source ports or source VLANs is copied to the RSPAN VLAN. This mirrored traffic is then forwarded over trunk ports to any destination session that is monitoring the RSPAN VLAN.

**Note** RSPAN does not require a dedicated reflector port to mirror traffic from either a source port or source VLAN.

## Link Aggregation Protocols

Fast EtherChannel (FEC) and Gigabit EtherChannel (GEC) are logically bundled physical interfaces that provide link redundancy and scalable bandwidth between network devices. Port Aggregation Protocol (PAgP) and Link Aggregation Control Protocol (LACP) help automatically create these channels by exchanging packets between Ethernet interfaces and negotiating a logical connection. PAgP is a

Cisco-proprietary protocol that can be run only on Cisco switches or on switches manufactured by vendors that are licensed to support PAgP. LACP is a standard protocol that allows Cisco switches to manage Ethernet channels between any switches that conform to the 802.3ad protocol. Because the CGESM supports both protocols, you can use either 802.3ad or PAgP to form port channels between Cisco switches.

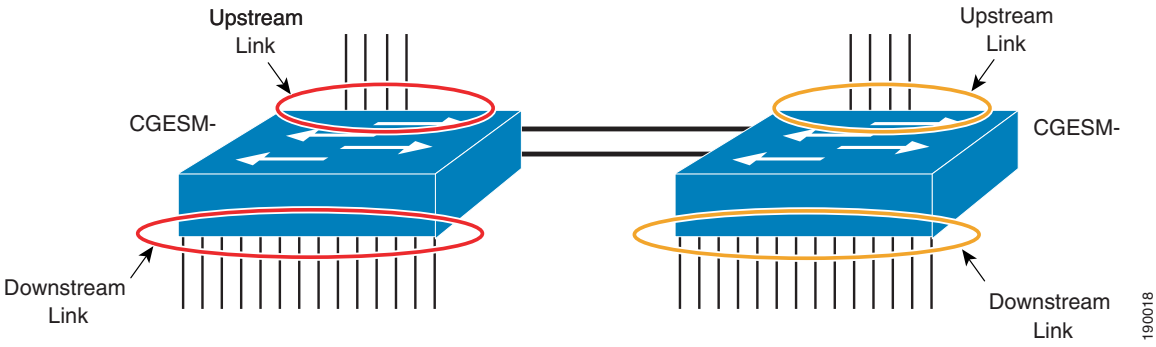
When using either of these protocols, a switch learns the identity of partners capable of supporting either PAgP or LACP and identifies the capabilities of each interface. The switch dynamically groups similarly configured interfaces into a single logical link, called a channel or aggregate port. The interface grouping is based on hardware, administrative, and port parameter attributes. For example, PAgP groups interfaces with the same speed, duplex mode, native VLAN, VLAN range, trunking status, and trunking type. After grouping the links into a port channel, PAgP adds the group to the spanning tree as a single switch port.

## Layer 2 Trunk Failover

Layer 2 Trunk failover is a high availability mechanism that allows the CGESM to track and bind the state of external interfaces with one or more internal interfaces. The four available Gigabit Ethernet uplink ports of the CGESM provide connectivity to the external network and are characterized as "upstream" links. The trunk failover feature may track these upstream interfaces individually or as a port channel. Trunk failover logically binds upstream links together to form a link state group. The internal interfaces of the CGESM provide blade server connectivity and are referred to as "downstream" interfaces in the trunk failover configuration. This feature creates a relationship between the two interface types where the link state of the "upstream" interfaces defined in a link state group determines the link state of the associated "downstream" interfaces.

Figure 2-16 illustrates the logical view of trunk failover on the CGESM. The two external port channels of Switch-1 and Switch-2 are configured as upstream connections in a link state group local to the switch. The 16 internal blade server ports are downstream interfaces associated with each local group.

***Figure 2-16 Trunk Failover Logical View***




Trunk failover places downstream devices into the same link state, "up" or "down", based on the condition of the link state group. If an uplink or upstream failure occurs, the trunk failover feature places the downstream ports associated with those upstream interfaces into a link "down" or inactive state. When upstream interfaces are recovered, the related downstream devices are placed in an "up" or active state. An average failover and recovery time for network designs implementing the trunk failover feature is three seconds.

Consider the following when configuring the trunk failover on the CGESM:

- Internal ports (Gigabit Ethernet 0/1–16) may not be configured as "upstream" interfaces.
- External ports (Gigabit Ethernet 0/19–24) may not be configured as "downstream" interfaces.

- Trunk failover does not consider STP. The state of the upstream connections determines the status of the link state group, not the STP state forwarding, blocking, and so on.
- Trunk failover of port channels requires that all of the individual ports of the channel fail before a trunk failover event is triggered.
- Interfaces cannot be members of more than one link state group.
- Do not configure an Etherchannel as a downstream interface.
- SPAN/RSPAN destination ports are automatically removed from the trunk failover link state groups.
- The CGESM is capable of defining two link state groups. This flexibility allows the administrator to define two different failover conditions for the downstream blade servers. This may be necessary when a server is not using NIC teaming, or two different uplink paths are defined on the switch.

# Using the HP BladeSystem p-Class Enclosure in the Data Center Architecture

The HP BladeSystem p-Class enclosure supports a maximum of two internal CGESM enhanced Layer 2 switches. Each switch provides 16 internal Gigabit Ethernet ports to support the blade servers within the enclosure. The HP BladeSystem p-Class supports up to eight blade servers (see Figure 2-13), each having multiple ProLiant NC series NICs (see Table 2-2). Note that the enclosure with enhanced backplane supports up to 16 blade servers (see Figure 2-15). Figure 2-17 and Figure 2-18 illustrate the physical layout of the chassis. The two interconnect bays house the CGESM switches that are connected via two 10/100/1000 cross connects on the backplane. Each switch also has separate dual-backplane connections to the individual server blade bays. This indicates that each server blade is dual-homed to the two internal switches.

***Figure 2-17 HP BladeSystem p-Class Enclosure with Proliant BL20p G3 Servers***

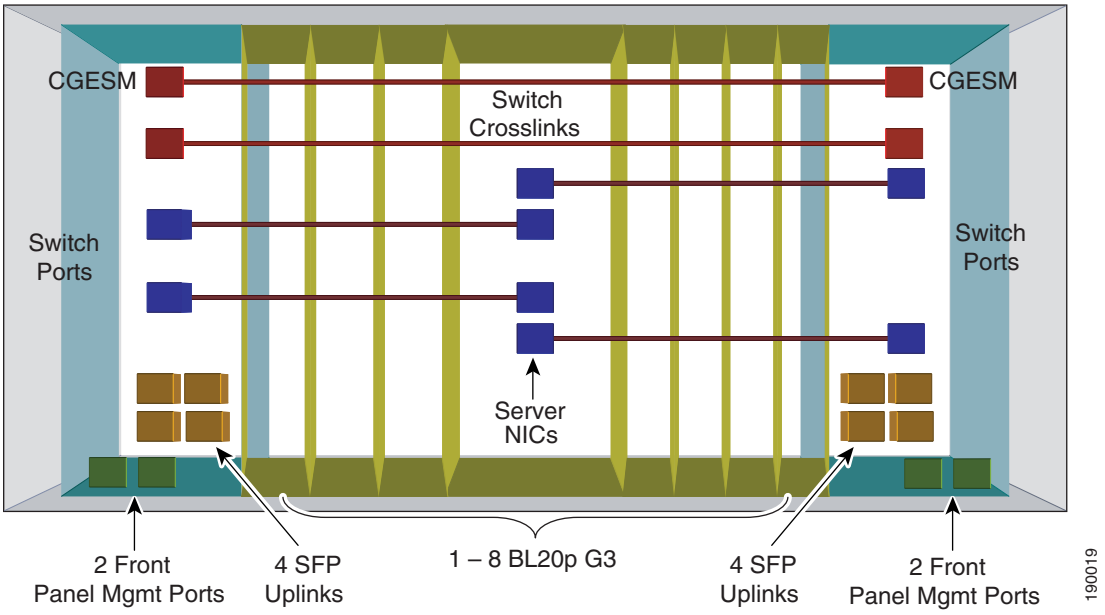
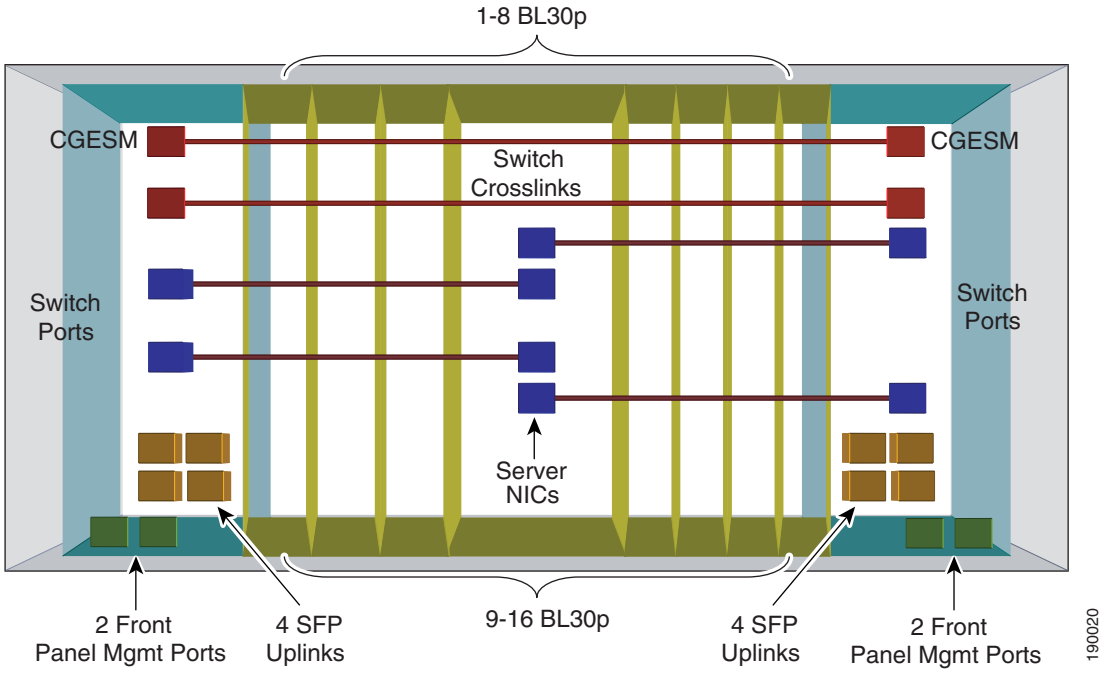


*Figure 2-18 HP BladeSystem p-Class Enclosure with Proliant BL30p Servers*




Each CGESM has four external SFP ports supporting 1000Base-SX and 10/100/1000Base-T on the rear of the enclosure and two external 10/100/1000BaseT ports on the front panel. These six ports provide connectivity to the data center or other external network. For more information, see the HP Blade Systems at the following URL:
http://h71028.www7.hp.com/enterprise/cache/80632-0-0-0-121.aspx#Servers

This section describes the design goals when deploying blade servers and the functionality supported by the CGESM in data centers. It includes the following topics:

- High Availability
- Scalability
- Management

## High Availability

Data centers are the repository of critical business applications that support the continual operation of an enterprise. Some of these applications must be accessible throughout the working day during peak times, and others at all times. The infrastructure of the data center, network devices, and servers must address these diverse requirements. The network infrastructure provides device and link redundancy combined with a deterministic topology design to achieve application availability requirements. Servers are typically configured with multiple NIC cards and dual-homed to the access layer switches to provide backup connectivity to the business application.

High availability is an important design consideration in the data center. An HP BladeSystem p-Class, using the CGESM, has a number of features and characteristics that contribute to a reliable, highly available network.

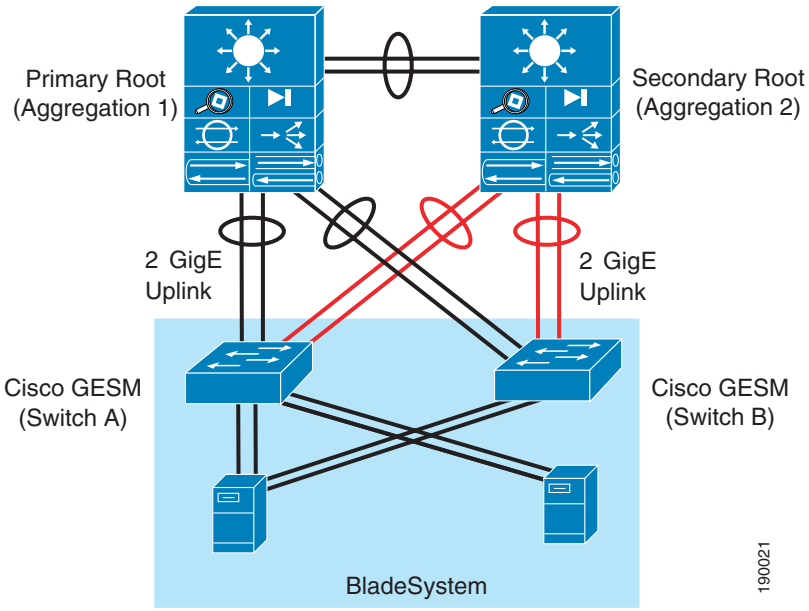
## High Availability for the Blade Enclosure Switching Infrastructure

High availability between the HP BladeSystem p-Class CGESMs and the aggregation layer switches requires link redundancy. Each CGESM in the HP BladeSystem p-Class uses four SFP uplinks for connectivity to the external network, which allows for redundant paths using two links each for more redundancy. Redundant paths implemented between the HP BladeSystem p-Class and each aggregation layer switch when each path uses two links provide a highly resilient design. However, this introduces the possibility of Layer 2 loops; therefore, a mechanism is required to manage the physical topology. The implementation of Rapid Spanning Tree Protocol (RSTP) ensures a fast converging, predictable Layer 2 domain between the aggregation layer and access switches (the CGESMs) when redundant paths are present.

The recommended design is a triangle topology, which delivers a highly available environment through redundant links and a spanning tree. It allows for multiple switch or link failures without compromising the availability of the data center applications.

As shown in Figure 2-19, each CGESM switch has two direct port channel connections to the Layer 2/Layer 3 aggregation layer where the primary STP root switch resides.

***Figure 2-19 Blade Enclosure Redundant Topology***




These channels support the publicly available subnets in the data center and traffic between servers. The server-to-server traffic that uses these uplinks is logically segmented through VLANs and may take advantage of network services available in the aggregation layer. There is also a port channel defined between the two blade enclosure switches. This path provides intra-chassis connectivity between the servers for VLANs defined locally on the blade enclosure switches. Clustering applications that require Layer 2 communication may take advantage of this traffic path, as well as mirrored traffic. Each of these port channels are composed of two-Gigabit Ethernet ports.

Cisco recommends using RPVST+ as the method for controlling the Layer 2 domain because of its predictable behavior and fast convergence. A meshed topology combined with RPVST+ allows only one active link from each blade enclosure switch to the root of the spanning tree domain. This design creates a highly available server farm through controlled traffic paths and the rapid convergence of the spanning tree.

The details of the recommended design are discussed in a later section.

### High Availability for the Blade Servers

The HP BladeSystem p-Class enclosure provides high availability to blade servers by multi-homing each server to the CGESMs. The two CGESMs housed in the interconnect bays are connected to the blade server over the backplane. Four backplane Gigabit Ethernet connections are available to every blade-server slot.

Multi-homing the server blades allows the use of network adapter (NIC) teaming driver, which provides another high availability mechanism to failover and load balance at the server level. The ProLiant NC series NICs support three modes of teaming:

- Network Fault Tolerance (NFT)—Creates a virtual interface by grouping the blade server network adapters into a team. One adapter is the primary active interface and all other adapters are in a standby state. The virtual adapter uses a single MAC address and a single Layer 3 address. NFT provides adapter fault tolerance by monitoring the state of each team member network connection. The standby NICs become active only if the primary NIC loses connectivity to the network.
- Transmit Load Balancing (TLB)—Supports adapter fault tolerance (NFT) and adds more functionality in the server for load balancing egress (transmit) traffic across the team. Note that a TLB team uses only one NIC to receive traffic. The load balancing algorithm is based on either the destination MAC or IP address. This teaming method provides better use of the bandwidth available for egress traffic in the network than NFT.
- Switch Assisted Load Balancing (SLB)—Extends the functionality of TLB by allowing the team to receive load balanced traffic from the network. This requires that the switch can load balance the traffic across the ports connected to the server NIC team. The CGESM supports the IEEE 802.3ad standard and Gigabit port channels.

For more information about the ProLiant NIC teaming features, see the following URL:
http://h18000.www1.hp.com/products/servers/networking/whitepapers.html

Layer 2 Trunk Failover combined with NIC teaming provides a complete high availability solution in a blade server environment. Trunk Failover allows teamed NICs to converge by disabling downstream server ports when upstream network failures are detected. This systematic approach makes the dual-homed architecture of the HP BladeSystem even more robust.

## Scalability

The ability of the data center to adapt to increased demands without compromising its availability is a key design consideration. The aggregation layer infrastructure and the services it provides must accommodate future growth in the number of servers or subnets it supports.

When deploying blade servers in the data center, there are the following two primary factors to consider:

- Number of physical ports in the aggregation and access layers
- Number of slots in the aggregation layer switches

### Physical Port Count

The introduction of blade systems into the data center requires greater port density at the aggregation layer. Blade systems, deployed with internal switches, provide their own access layer. The cabling and maximum number of servers per enclosure is predetermined. Scaling the aggregation layer ports to accommodate the blade system uplinks is an area that requires attention.

As shown in Figure 2-19, each CGESM requires four Gigabit Ethernet ports from the aggregation layer switches. The number of physical ports that an aggregation-layer switch can support equals the number of ports per slot times the number of available slots.

It is important to remember that aggregation switches provide data center services such as load balancing, security, and network analysis that may require dedicated ports for appliances or slots for integrated services. This directly affects the number of ports available for access layer connectivity.

Table 2-3 lists the number of blade systems supported by a single line card with varying port counts. This table is based on the recommended topology shown in Figure 2-4, where each blade system is dual-homed to the aggregation layer over two Gigabit Ethernet port channels.

***Table 2-3 Blade System Support per Aggregate Switch Line Card***

| Type of Line Card | Uplinks per CGESM | Total Uplinks /Blade System Enclosure (Two CGESM/Enclosure) | Blade Systems per Line Card |
|---|---|---|---|
| 8-port Gigabit Ethernet | 2 | 4 | 2 |
| 16-port Gigabit Ethernet | 2 | 4 | 4 |
| 48-port Gigabit Ethernet | 2 | 4 | 12 |

Table 2-3 highlights the finite number of BladeSystems supported by a single aggregate switch line card. This table implies that the aggregate layer must provide line card density for a scalable BladeSystem environment.

## Slot Count

The data center infrastructure must be flexible enough to allow growth both in server capacity and service performance. Connecting a blade system directly into the aggregation layer places more significance on the number of slots available to accommodate blade system uplinks and integrated services.

Traditionally, the access layer provides the port density necessary to allow the physical growth of server farms. Modular access layer switches offer connectivity to densely packed server farms over a few uplinks. The aggregation layer switches support a limited number of uplinks from the access layer. With this model, the number of servers supported per uplink is high.

Blade systems use more aggregation layer resources per server than this traditional deployment model. Each uplink from a blade enclosure provides connectivity to a maximum of 16 servers. The aggregation layer must be flexible enough to manage the increased demand for ports and slots in this blade server system environment.

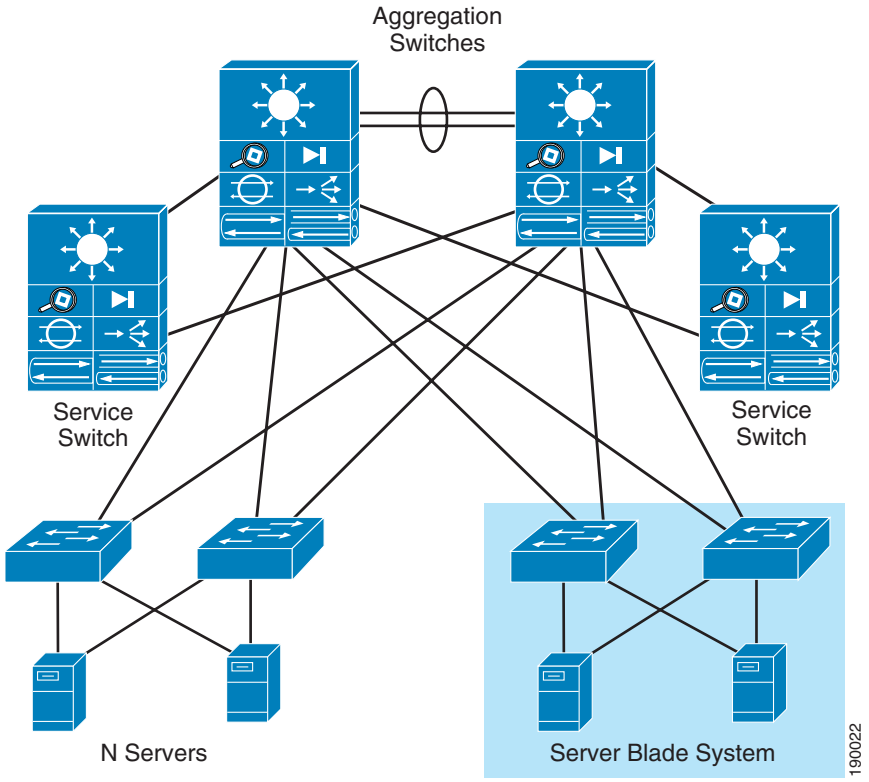
To scale the server farm, use an aggregation layer switch that provides an ample number of slots for line cards and/or service module expansion.

In addition, consider using the following two options (which are not mutually exclusive):

- Deploying service switches in the aggregation layer (as shown in Figure 2-20)
- Using a data center core to accommodate multiple aggregation layer modules

Service switches are deployed in the aggregation layer to host integrated data center services such as load balancing, intrusion detection, and network analysis. Relocating these services to a separate switch frees ports and slots in the aggregation layer switches. This design allows the aggregation switches to commit more slots and ultimately, more ports to the Layer 2 connectivity of the server farms. Figure 2-20 shows a service switch deployment.

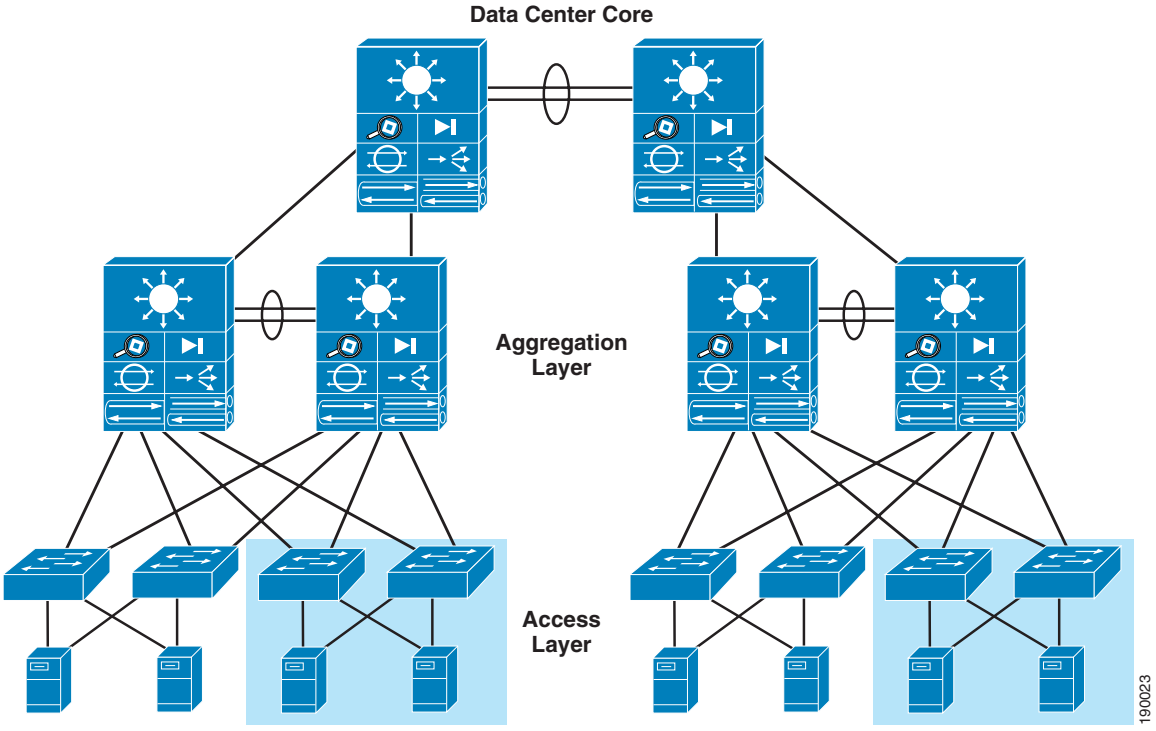
***Figure 2-20 Data Center Scaling with Service Switches***




The data center core is a mechanism to replicate and horizontally scale the data center environment. In the recommended design, the aggregation and access layer is regarded as a module that can be duplicated to extend the enterprise. Each data center module provides its own network services locally in the aggregation switches. This approach allows the network administrator to determine the limits of each data center module and to replicate as necessary.

Figure 2-21 shows the data center core design. The aggregation switches for each data center module are Layer 3-attached to the core. In addition, the aggregation switches house the service modules required to support the server farms.

*Figure 2-21 Data Center Core Design*




## Management

The CGESM is accessible for management and configuration by any of the following traffic paths:

- Out-of-band management
- In-band management
- Serial/console port

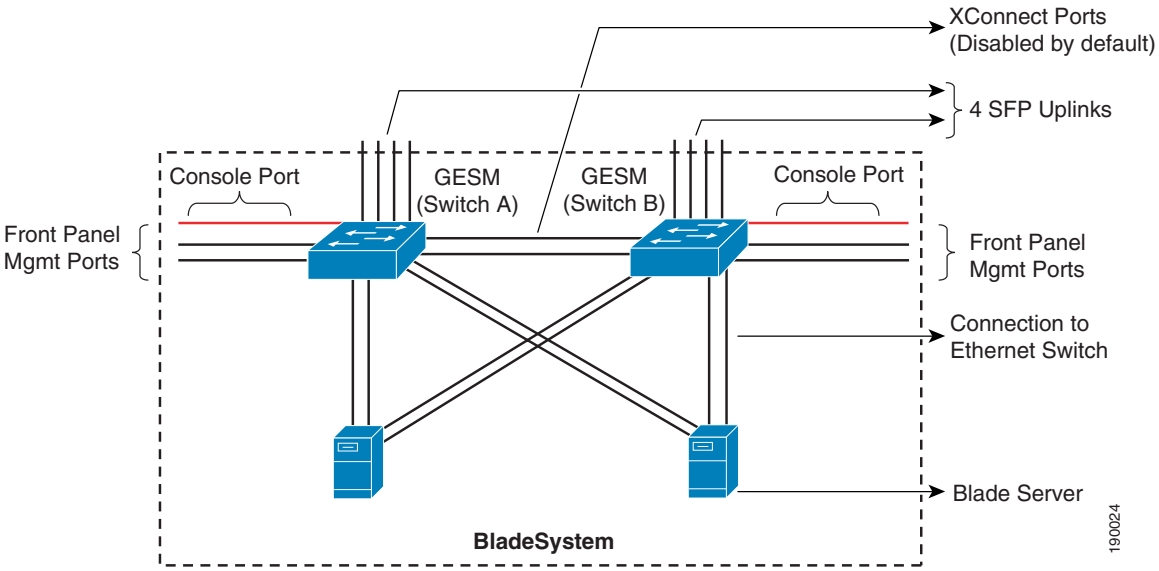
These traffic paths provide three management options for network administration and support various user and application interfaces to the CGESM. The remote management of the blade servers within the HP BladeSystem p-Class enclosure is critical to an efficient and scalable data center. The iLO connectivity options provided via the enclosure to the blade servers are also discussed.

### Out-of-Band Management

Out-of-band management is the practice of dedicating an interface on the managed device for carrying management traffic. It is also the recommended management method for HP BladeSystems. Out-of-band management isolates the management and data traffic and provides a more secure environment.

Figure 2-22 illustrates the interfaces available for connectivity. Cisco recommends using the front panel ports for connectivity to the management domain.

*Figure 2-22 Blade Enclosure*




The CGESM allows only one switched virtual interface (SVI) to be active. By default, the SVI is created as VLAN 1 and it is disabled in an administratively down state. Cisco recommends that a VLAN other than VLAN 1 be used as the management VLAN. Therefore, it is important to create an SVI with another VLAN and to allow this VLAN on the external front panel ports.

For best practices in selecting the management VLAN, see the following URL: http://www.cisco.com/en/US/products/hw/switches/ps700/products_white_paper09186a00801b49a4.shtml

## In-Band Management

In-band management uses logical isolation to separate management traffic from data traffic. VLANs segregate the two traffic types that are sharing the bandwidth of the uplink ports. This practice is common where applications running on the servers must be managed along with the network infrastructure devices.

In-band management traffic uses the uplink trunk ports located on the rear of the CGESMs for management. Cisco recommends using a VLAN other than VLAN 1 for management.

## Serial/Console Port

The front panel of the CGESM has a single RJ-45 serial port that can be used to manage the switch through the command-line interface (CLI). The CLI can be accessed by connecting directly to the console port with the serial port of a workstation or remotely by using terminal servers and IP connectivity protocols such as Telnet.

## Management Options

The CGESM switch is manageable through the following methods:

- HTTP-based device manager GUI
- SNMP-based management applications
- Cisco IOS software CLI

The embedded device manager on the CGESM provides a GUI interface to configure and monitor the switch through a web browser. This requires using either in-band or out-of-band management and enabling the HTTP/HTTPS server on the switch. The HTTP server and SSL are enabled by default.

SNMP-compatible management utilities are supported through a comprehensive set of MIB extensions and through four remote monitoring (RMON) groups. CiscoWorks2000 and HP OpenView are two such management applications. SNMP versions 1, 2, and 3 are available on the switch (Cisco IOS software crypto image).

The CLI delivers the standard Cisco IOS software interface over Telnet or the console port. The use of SSH for CLI access is recommended.

**Note** For more information about the embedded device manger, see the online help on the switch CLI.

For more information about Cisco MIBs, see the following URL:
http://www.cisco.com/public/sw-center/netmgmt/cmtk/mibs.shtml

For more information about the management options for the HP BladeSystem, see the following URL:
http://h18004.www1.hp.com/products/blades/components/management.html

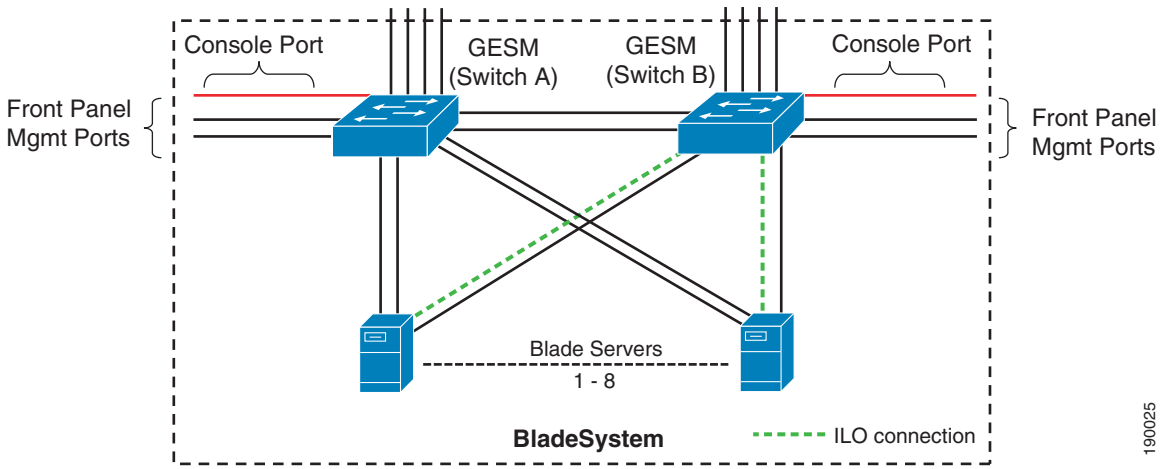
## HP BladeSystem p-Class iLO Connectivity

The iLO provides remote management capabilities and is standard with BL p-Class server blades. Remote power, console, and diagnostics are just a few of the advanced functions iLO provides. Table 2-2 shows that each of the blade servers supports a dedicated iLO NIC. The HP BladeSystem p-Class enclosure provides two methods to access this management interface:

- Through the CGESM located in interconnect bay B
- Through an enhanced BladeSystem enclosure

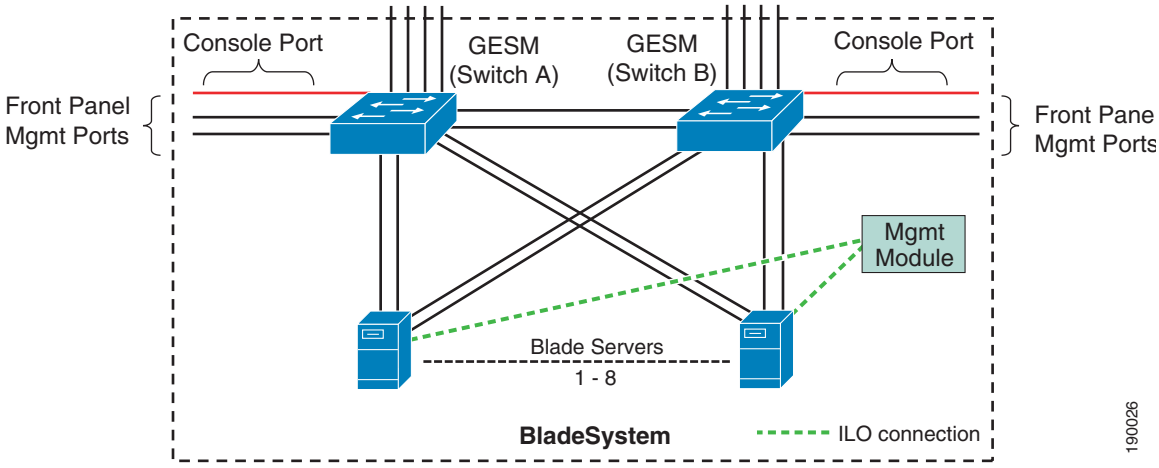
An HP BladeSystem p-Class enclosure without the enhanced backplane features provides connectivity to the dedicated iLO NIC through the CGESM located in interconnect bay B. The iLO NIC auto-negotiates to 100 Mbps and uses one of the CGESM ports assigned to that server bay. Figure 2-23 illustrates the iLO interface connection in the absence of an enhanced backplane.

***Figure 2-23 HP BladeSystem p-Class iLO Connectivity***



The enhanced backplane of the HP BladeSystem p-Class enclosure allows each server to use a dedicated Ethernet port for iLO connectivity. As shown in Figure 2-24, the iLO connection is independent of the CGESM. The blade server management module located on the rear of the chassis provides access to each of the iLO interfaces through a single RJ-45 cable.

*Figure 2-24 HP BladeSystem pClass with Enhanced Backplane iLO Connectivity*




**Note** The Proliant BL30p server blade requires the use of an enhanced backplane.

# Design and Implementation Details

- Network Management Recommendations
- Network Topologies using the CGESM
- Layer 2 Looped Access Layer Design—Classic "V"
- Layer 2 Looped Access Layer Design—"Square"
- Layer 2 Loop-Free Access Layer Design—Inverted "U"
- Configuration Details

## Network Management Recommendations

An out-of-band (OOB) network is recommended for managing the CGESM switch. OOB management provides an isolated environment for monitoring and configuring the switch. Isolation is achieved by deploying a physically separate management network or by logically separating the traffic with management VLANs.

The CGESM switch has two external Gigabit Ethernet ports located at the front of the chassis that may be used to support network monitoring devices and network management traffic. The use of secure protocols, such as SSH or HTTPS, maintains the integrity of communications between the switch and the management station. The console port positioned at the front of the CGESM is another option for connectivity to the OOB network.

## Network Topologies using the CGESM

The following physical topologies are discussed in this section:

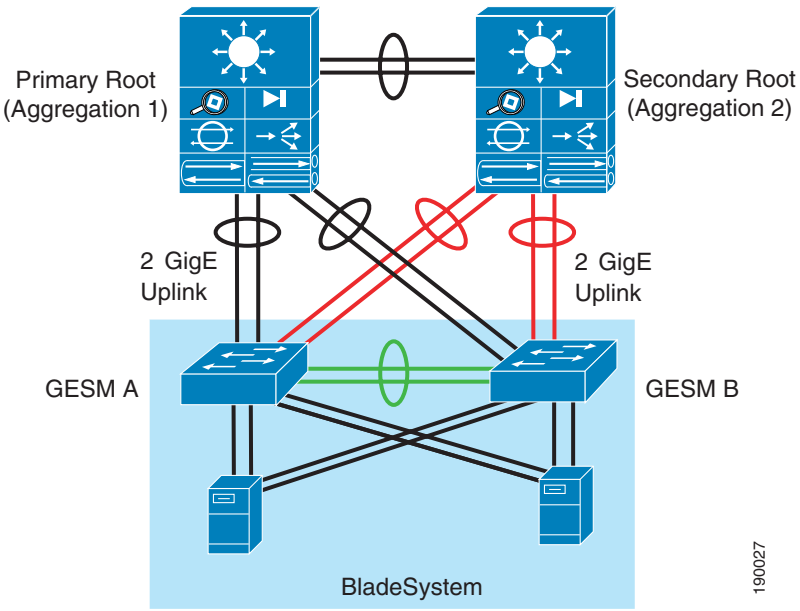
- Layer 2 Looped Access Layer Design—Classic "V"
- Layer 2 Looped Access Layer Design—"Square"
- Layer 2 Loop-Free Access Layer Design—Inverted "U"

These network designs emphasizes high availability in the data center by eliminating any single point of failure, and by providing deterministic traffic patterns, and predictable behavior during times of network convergence. The configuration example included uses a pair of Cisco Catalyst 6513 as the aggregation layer platform. This Layer 2/Layer 3 switching platform supports slot density and integrated network services required by data centers deploying blade systems. An HP BladeSystem p-Class server blade enclosure with two CGESMs composes the Layer 2 access layer.

## Layer 2 Looped Access Layer Design—Classic "V"

Figure 2-25 shows a blade system deployment in the data center using the classic triangle topology. There is no single point of failure in this deployment model. The CGESMs are dual-homed to the aggregation layer, providing link redundancy. STP manages the physical loops created by the uplinks between the aggregation and access switches, assuring a predictable and fast converging topology. In Figure 2-25, the black links are in spanning tree forwarding state and the red links are in blocking state. The green links represent the internal cross connects that are disabled by default.

***Figure 2-25 Recommended Topology HP BladeSystem p-Class with CGESMs***




RPVST+ fulfills the high availability requirements of this design and is the recommended mode of spanning tree operation. RPVST+ provides fast convergence (less than one second) in device or uplink failure scenarios. In addition, RPVST+ offers enhanced Layer 2 features for the access layer with integrated capabilities equivalent to UplinkFast and BackboneFast.

The connection between the two internal blade switches in Figure 2-25 supports local traffic limited to the HP BladeSystem; for example, clustering applications, or management traffic such as remotely mirrored (RSPAN) traffic. This connection does not carry a publicly accessible subnet (for example, a VLAN that exists on the uplinks to the aggregation switches). If this were the case, another set of interfaces would have to be accounted for in the STP calculations. Therefore, to create a less complex STP domain, these cross-connect interfaces are removed from the equation by clearing the public VLANs from the links.

The HP BladeSystem p-Class server blade NICs support the logical separation of VLANs via trunking. This allows each NIC to accommodate the public and the private VLANs on the CGESMs. In addition, servers such as the BL20P G3 series are dual-homed to each of the two CGESMs in the HP BladeSystem enclosure (see Figure 2-19). This structural design allows for the physical separation of public and private VLANs between two NICs homed to the same CGESM.

A series of network convergence tests were performed to verify and characterize the high availability features of the recommended design. These tests consisted of passing traffic between an external client device and the blade servers while monitoring packet loss. The following test cases were used:

- Uplink failure and recovery between Switch-A and the primary root
- Uplink failure and recovery between Switch-B and the primary root
- Switch-A failure and recovery
- Switch-B failure and recovery
- Primary root switch failure and recovery
- Secondary root switch failure and recovery

These tests revealed the intricacies of fast convergence in the data center and the necessity for a holistic approach to high availability.

Test cases that did not involve the failure of the active HSRP aggregation switch resulted in an average failover time of about one second. Failing the active HSRP device requires convergence at Layer 3 and resulted in a recovery time that reflected the settings of the HSRP timers.

It is possible to tune the HSRP timers for sub-second convergence. However, when multiple HSRP devices are involved, the recovery time is typically in the five-second range.

In this topology, two Gigabit Ethernet links comprise the port channel uplinks between the access and aggregation layers. This configuration allows a single link to fail without triggering STP convergence.

**Note** The default gateway for the servers is the HSRP address of the Layer 3 aggregation switches. Failover times may be affected if the default gateway of the server is located on another device, such as a load balancer or firewall.

The recommended topology provides a high level of availability to the blade servers except in one failure scenario. If both the uplinks to each of the aggregation switches from a single CGESM are unavailable, the server NICs homed to that CGESM are not notified. The blade servers are unaware of the disconnection between the access layer switches (CGESMs) and the aggregation layer switches and continue to forward traffic. To address this breakdown in network connectivity, use one of the following methods:

- Use the NIC teaming features of the ProLiant blade servers with Layer 2 Trunk Failover.
- Deploy load balancing in front of the blade server farm.

In addition, the NIC teaming features of the ProLiant blade servers provide redundancy at the network adapter level. Stagger the preferred primary NICs between the two Cisco switches in the enclosure to increase server availability. Assigning the primary NIC is a straightforward process. The NIC teaming

software provides a GUI interface or a small configuration file, depending on the operating system, to construct the team. HP also offers network-aware teaming software to verify and detect network routes. For more information about these features, see the ProLiant Essential Intelligent Network Pack at the following URL: http://h18004.www1.hp.com/products/servers/proliantessentials/inp/index.html

The ProLiant blade server NIC teaming functionality combined with the Cisco Layer 2 Trunk Failover feature provides a comprehensive high availability solution to the blade servers. On detection of an upstream link failure from the CGESM, the associated downstream server ports are disabled by the CGESM, which allows the NIC team to redirect traffic over the remaining active NICs in the team homed to another CGESM. Multiple link state groups may be defined on the CGESM to allow for redundant uplink paths.

By monitoring the health of a server farm, a load balancer can bypass the network failure by redirecting traffic to available servers. This helps ensure fulfillment of end user requests despite the network failure.

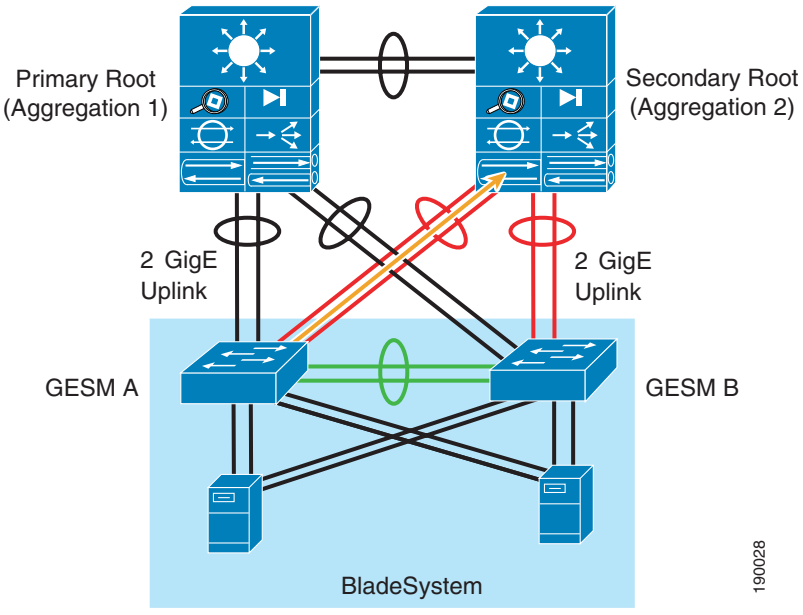
The recommended network topology illustrated in Figure 2-25 allows for traffic monitoring either locally or remotely using Switched Port Analyzer (SPAN). Local SPAN supports monitoring of network traffic within one switch, while remote SPAN (RSPAN) allows the destination of mirrored traffic to be another switch within the data center. The source of mirrored traffic for a SPAN or RSPAN session may be one or more ports or VLANs.

Local SPAN is readily supported by the CGESM over one of the two external Gigabit Ethernet ports located on the front panel of the switch. This RJ-45 connection is an ideal location to attach intrusion detection or other network analysis device.

RSPAN requires a VLAN to carry the mirrored traffic to the remote destination switch. In the recommended topology, the secondary aggregation switch is the RSPAN destination, where an analysis device, such as the integrated Network Analysis Module (NAM), resides.

Figure 2-26 illustrates the traffic path of the RSPAN VLAN. The RSPAN VLAN uses the uplink between the blade switch and the secondary aggregation switch. This uplink is blocking under normal conditions for regular VLANs. As a result, bandwidth utilization is only a concern when the uplink is forwarding and sharing the path with production traffic.

***Figure 2-26 RSPAN Traffic Path***



### Configuring the Aggregate Switches

Complete the following sequence of steps on the aggregate switches:

1. VLAN configuration
2. RPVST+ configuration
3. Primary and secondary root configuration
4. Configuration of port channels between aggregate switches
5. Configuration of port channels between aggregate and CGESM switches
6. Trunking the port channels between aggregate switches
7. Configuration of default gateway for each VLAN

**Note** Configuration Details, page 2-52 describes each of these steps.

### Configuring the CGESM Switches

Complete the following sequence of steps on the CGESM switches:

1. VLAN configuration
2. RPVST+ configuration
3. Configuration of port channels between the CGESM and aggregate switches
4. Trunking port channels between the CGESM and aggregate switches
5. Configuration of server ports on the CGESM
6. Configure Layer 2 Trunk Failover

**Note** Configuration Details, page 2-52 describes each of these steps.

### Additional Aggregation Switch Configuration

The following recommendations help integrate the CGESM switches into the data center:

1. Enable Root Guard on the aggregate switches links connected to the switches in the blade enclosure.

   The spanning tree topology is calculated, and one of the primary parameters involved in this equation is the location of the root switch. Determining the position of the root switch in the network allows the network administrator to create an optimized forwarding path for traffic. Root Guard is a feature designed to control the location of the root switch.

   The aggregation switches should employ the **spanning-tree guard root** command on the port channel interfaces connected to the blade switches.

2. Allow only those VLANs that are necessary on the port channel between the aggregate switch and the blade switches.

   Use the **switchport trunk allowed vlan** *vlanID* command to configure the port channel interfaces of the aggregate switch to allow only those VLANs indicated with the *vlanID* option.

### Additional CGESM Configuration

1. Enable BPDU Guard on the internal server ports of the switch

Use the **spanning-tree bpduguard enable** command to shut down a port that receives a BPDU when it should not be participating in the spanning tree.

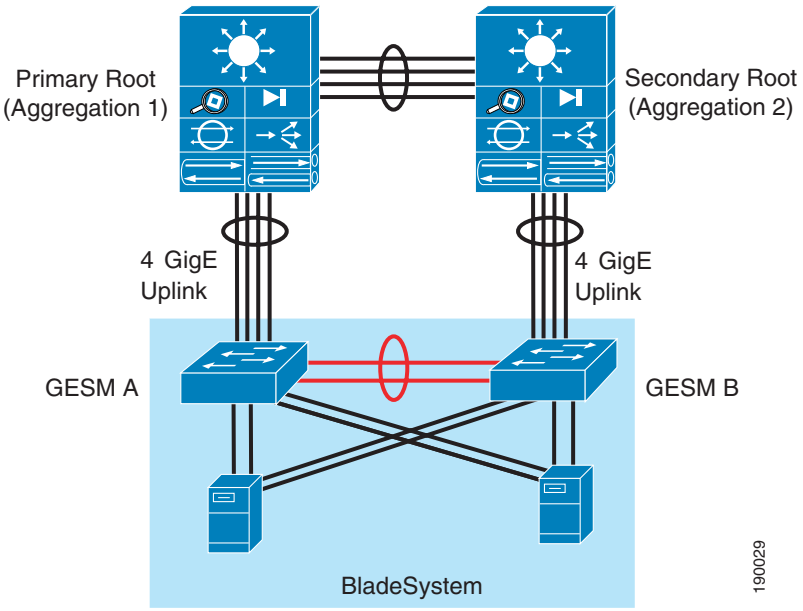
2. Allow only those VLANs that are necessary on the port channels between the aggregate switches and the blade switches.

   Use the **switchport trunk allowed vlan** *vlanID* command to configure the port channel interfaces of the switch to allow only those VLANs indicated with the *vlanID* option.

## Layer 2 Looped Access Layer Design—"Square"

Figure 2-27 illustrates an alternative topology that relies on RPVST+ to account for redundant paths in the network. The two aggregate switches connect to each other via a port channel supporting the server farm VLANs. The four external uplinks of each CGESM are channeled and connected to one of the two aggregate switches. The internal connections between the two CGESMs complete the loop.

***Figure 2-27 Alternate Topology HP BladeSystem p-Class with CGESMs***




This design uses the links between the two CGESMs as a redundant path for blade server traffic. In Figure 2-27, the black links are in spanning tree forwarding state and the red links are in blocking state. These links are in blocking state by default. The use of a longer path cost value provides for a more granular calculation of the topology based on the available link bandwidth (see Cisco IGESM Features, page 2-3). This feature is enabled with the **spanning-tree pathcost method long** command. RPVST+ should be used in this network design for its fast convergence and predictable behavior.

**Note** This design uses a lower bandwidth path when an uplink failure occurs on either CGESM A or CGESM B. To increase the bandwidth of the redundant path between the CGESMs, consider using the external ports of CGESM A and B in the EtherChannel.

The following convergence tests were conducted with this alternate topology:

- Uplink failure and recovery between Switch-A and the primary root

- Uplink failure and recovery between Switch-B and the secondary root
- Failure and recovery of Switch-A and Switch-B
- Failure and recovery of the primary and secondary root switches

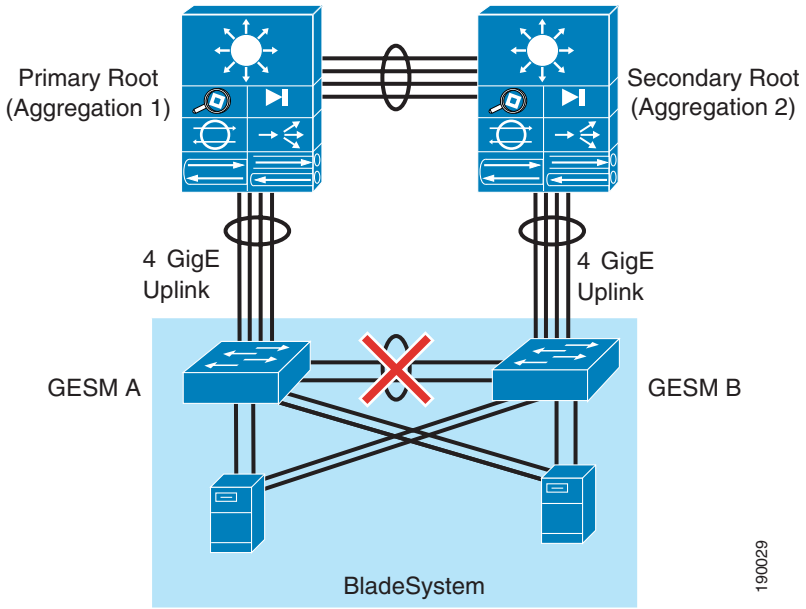
These tests yielded results similar to the recommended topology. Layer 2 convergence occurs in approximately one second. As stated previously, recovery at Layer 3 is dependent on the HSRP settings of the aggregate switches (see Network Management Recommendations, page 2-45). In the Cisco test bed, the failure of the active HSRP device typically increased the convergence time to five seconds.

The design in Figure 2-27 supports traffic monitoring via SPAN and/or RSPAN. For example, a network analysis device connected to the external ports on the front of the CGESM may capture locally mirrored traffic. Alternatively, RSPAN traffic may be carried on the CGESM uplinks if bandwidth utilization is not a concern. For the steps to configure traffic monitoring, see Configuration Details, page 2-52.

## Layer 2 Loop-Free Access Layer Design—Inverted "U"

Figure 2-28 shows a Layer 2 loop-free access layer design commonly referred to as an inverted "U". The two aggregate switches connect to each other via a port channel supporting the server farm VLANs. The four external uplinks of each CGESM are channeled and connected to one of the two aggregate switches. The internal connections between the two CGESMs remain disabled (the default state of these links) to disrupt the loop. This design requires that the connections between the two CGESMs remain disabled to prevent a loop condition. Cisco recommends that RPVST+ be enabled to account for inadvertent loops created through human errors during configuration or uplink wiring.

***Figure 2-28 Layer 2 Loop-Free Access Layer Design—Inverted "U"***




The inverted "U" design provides a highly available blade server environment using NIC teaming and Layer 2 trunk failover. The CGESM Trunk Failover feature disables downstream server ports when the uplink port channel fails. Disabling the downstream server ports of the CGESM allows a properly configured ProLiant blade server NIC team to converge traffic to the remaining active NICs homed to the other CGESM in the BladeSystem.

Convergence tests with this topology revealed that approximately three seconds of downtime is experienced when uplink failures occur on the CGESM with the active server NICs. The CGESM, employing Layer 2 Trunk Failover, properly disabled downstream server ports when the uplink failure condition occurred.

### Configuring the Aggregate Switches

Complete the following sequence of steps on the aggregate switches:

1. VLAN configuration
2. RPVST+ configuration
3. Primary and secondary root configuration
4. Configuration of port channels between aggregate switches
5. Configuration of port channels between aggregate and CGESM switches
6. Trunking the port channels between aggregate switches
7. Configuration of default gateway for each VLAN

**Note** Configuration Details, page 2-52 describes each of these steps.

### Configuring the CGESM Switches

Complete the following sequence of steps on the CGESM switches:

1. VLAN configuration
2. RPVST+ configuration
3. Configuration of port channels between the CGESM and aggregate switches
4. Trunking port channels between the CGESM and aggregate switches
5. Configuration of server ports on the CGESM
6. Configure Layer 2 Trunk Failover

**Note** Configuration Details, page 2-52 describes each of these steps.

## Configuration Details

This section describes the configuration steps required for implementing the topologies discussed in this guide. The configuration for the following are discussed:

- VLAN
- RPVST+
- Inter-Switch Link
- Port Channel
- Trunking
- Server Port
- Server Default Gateway

- RSPAN
- Layer 2 Trunk Failover

## VLAN

To configure the VLANs on the switches, complete the following tasks:

**Step 1** Set the VLAN trunking protocol administrative domain name and mode as follows:

```
(config)# vtp domain <domain name>
(config)# vtp mode transparent
```

**Step 2** Configure the server farm VLANs as follows:

```
(config)# vlan 60
(config-vlan)# name bladeservers
(config-vlan)# state active
```

## RPVST+

**Step 1** Configure STP to manage the physical loops in the topology. Cisco recommends using RPVST+ for its fast convergence characteristics. Set the STP mode an each aggregation switch as follows:

```
(config)# spanning-tree mode rapid-pvst
```

**Step 2** Configure the path cost to use 32 bits in the STP calculations:

```
(config)# spanning-tree pathcost method long
```

**Step 3** Configure the root switch as follows:

```
(config)# spanning-tree vlan <vlan range> root primary
```

**Step 4** Configure the secondary root switch as follows:

```
(config)# spanning-tree vlan <vlan range> root secondary
```

## Inter-Switch Link

The topologies discussed in this guide require connectivity between the switches. The following three types of inter-switch connections exist:

- Aggregate-1 to Aggregate-2
- Aggregate-1 or Aggregate-2 to Blade Enclosure Switch-A or Switch-B
- HP BladeSystem Switch-A to Switch-B

Each of these connections are Layer 2 EtherChannels consisting of multiple physical interfaces bound together as a channel group or port channel. These point-to-point links between the switches should carry more than one VLAN; therefore, each is a trunk.

## Port Channel

Link Aggregate Control Protocol (LACP) is the IEEE standard for creating and managing EtherChannels between switches. Each aggregate switch uses this feature to create a port channel across the line cards. The use of multiple line cards within a single switch reduces the possibility of the point-to-point port channel becoming a single point of failure in the network.

**Step 1** Configure the active LACP members on Aggregate-1 to CGESM Switch-A as follows:

```
(config)# interface GigabitEthernet12/1
(config-if)# description <<** Connected to Switch-A **>>
(config-if)# channel-protocol lacp
(config-if)# channel-group 1 mode active
(config)# interface GigabitEthernet11/1
(config-if)# description <<** Connected to Switch-A **>>
(config-if)# channel-protocol lacp
(config-if)# channel-group 1 mode active
```

**Step 2** Configure the passive LACP members on CGESM Switch-A as follows:

```
(config) # interface GigabitEthernet0/19
(config-if)# description <<** Connected to Aggregation-1 **>>
(config-if)# channel-group 1 mode on
(config) # interface GigabitEthernet0/20
(config-if)# description <<** Connected to Aggregation-1 **>>
(config-if)# channel-group 1 mode on
```

## Trunking

Use the following guidelines when configuring trunks:

- Allow only those that are necessary on the trunk
- Use 802.1q trunking
- Tag all VLANs over a trunk from the aggregation switches

**Step 1** Configure trunks using the standard encapsulation method 802.1q as follows:

```
(config-if)# switchport trunk encapsulation dot1q
```

**Step 2** Define the VLANs permitted on a trunk as follows:

```
(config-if)# switchport trunk allowed vlan <VLAN IDs>
```

**Step 3** Modify the VLANs allowed on a trunk using one of the following commands:

```
(config-if)# switchport trunk allowed vlan add <VLAN IDs>
(config-if)# switchport trunk allowed vlan remove <VLAN IDs>
```

**Step 4** Define a port as a trunk port as follows:

```
(config-if)# switchport mode trunk
```

**Note** The auto-negotiation of a trunk requires that the ports be in the same VTP domain and be able to pass DTP frames.

**Step 5** To secure and enforce a spanning tree topology, configure the Root Guard feature on the aggregate switch interfaces that connect to the blade switches.

The following is an example of the interface configuration between the aggregate and blade switch with root guard enabled:

```
(config)# interface GigabitEthernet12/13
config-if)# description <text>
config-if)# no ip address
config-if)# switchport
config-if)# switchport trunk encapsulation dot1q
config-if)# switchport trunk native vlan <vlan id>
config-if)# switchport trunk allowed vlan <vlan id>
config-if)# switchport mode trunk
config-if)# spanning-tree guard root
config-if)# channel-protocol lacp
config-if)# channel-group <group id> mode active
```

## Server Port

A blade server is assigned a specific port on the blade switch. This is pre-determined by the physical slot the blade server occupies in the chassis. Table 2-3 correlates the server and switch ports.

*Table 2-4 Correlation of Server and Switch Ports*

| IOS CLI Identifier | Actual Port Location in 8-Slot Server Chassis | Actual Port Location in 16-Slot Server Chassis |
|---|---|---|
| GigabitEthernet 0/1 | Server Slot 1 | Server Slot 1 |
| GigabitEthernet 0/2 | Server Slot 1 | Server Slot 2 |
| GigabitEthernet 0/3 | Server Slot 2 | Server Slot 3 |
| GigabitEthernet 0/4 | Server Slot 2 | Server Slot 4 |
| GigabitEthernet 0/5 | Server Slot 3 | Server Slot 5 |
| GigabitEthernet 0/6 | Server Slot 3 | Server Slot 6 |
| GigabitEthernet 0/7 | Server Slot 4 | Server Slot 7 |
| GigabitEthernet 0/8 | Server Slot 4 | Server Slot 8 |
| GigabitEthernet 0/9 | Server Slot 5 | Server Slot 9 |
| GigabitEthernet 0/10 | Server Slot 5 | Server Slot 10 |
| GigabitEthernet 0/11 | Server Slot 6 | Server Slot 11 |
| GigabitEthernet 0/12 | Server Slot 6 | Server Slot 12 |
| GigabitEthernet 0/13 | Server Slot 7 | Server Slot 13 |
| GigabitEthernet 0/14 | Server Slot 7 | Server Slot 14 |
| GigabitEthernet 0/15 | Server Slot 8 | Server Slot 15 |
| GigabitEthernet 0/16 | Server Slot 8 | Server Slot 16 |
| GigabitEthernet 0/17 | Cross Connect Port 1 | Cross Connect Port 1 |
| GigabitEthernet 0/18 | Cross Connect Port 2 | Cross Connect Port 2 |
| GigabitEthernet 0/19 | SFP/Uplink Port 1 | SFP/Uplink Port 1 |
| GigabitEthernet 0/20 | SFP/Uplink Port 2 | SFP/Uplink Port 2 |
| GigabitEthernet 0/21 | SFP/Uplink Port 3 | SFP/Uplink Port 3 |
| GigabitEthernet 0/22 | SFP/Uplink Port 4 | SFP/Uplink Port 4 |

*Table 2-4 Correlation of Server and Switch Ports*

| | | |
|---|---|---|
| GigabitEthernet 0/23 | RJ45/Front Panel Port 1 | RJ45/Front Panel Port 1 |
| GigabitEthernet 0/24 | RJ45/Front Panel Port 2 | RJ45/Front Panel Port 2 |

The server ports on the blade switch support a single VLAN access and trunk configuration modes. The operational mode chosen should support the server NIC configuration (that is, a trunking NIC is attached to a trunking switch port). Enable PortFast for the edge devices.

The BPDU Guard feature disables a port that receives a BPDU. This feature protects the STP topology by preventing the blade server from receiving BPDUs. A port disabled via the BPDU Guard feature must be recovered by an administrator manually. Enable the BPDU Guard feature on all server ports that should not be receiving BPDUs.

Port Security limits the number of MAC addresses permitted to access the blade switch port. Configure the maximum number of MAC addresses expected on the port.

**Note** The NIC teaming driver configuration (that is, the use of a virtual MAC address) must be considered when configuring Port Security.

```
interface GigabitEthernet0/1
 description <<** BladeServer-1 **>>
 switchport trunk encapsulation dot1q
 switchport trunk allowed vlan 10,60
 switchport mode trunk
 switchport port-security aging time 20
 switchport port-security maximum 1 vlan 10,60
 no cdp enable
 spanning-tree portfast trunk
 spanning-tree bpduguard enable
end
```

## Server Default Gateway

The default gateway for a server is a Layer 3 device located in the aggregation layer of the data center. This device may be a firewall, a load balancer, or a router. Using protocols such as HSRP protect the gateway from being a single point of failure and create a highly available date center network. HSRP allows the two aggregate switches to act as a single virtual router by sharing a common MAC and IP address between them. Define a switched VLAN interface on each aggregate switch and use the HSRP address as the default gateway of the server farm.

**Step 1** Configure Aggregation-1 as the active HSRP router. The **priority** command helps to select this router as the active router because it has a greater value.

```
interface Vlan10
 description <<**  BladeServerFarm – Active  **>>
 ip address 10.10.10.2 255.255.255.0
 no ip redirects
 no ip proxy-arp
 arp timeout 200
 standby 1 ip 10.10.10.1
 standby 1 timers 1 3
 standby 1 priority 51
 standby 1 preempt delay minimum 60
 standby 1 authentication <password>
end
```

**Step 2** Configure Aggregation-2 as the standby HSRP router as follows:

```
interface Vlan10
 description <<**  BladeServerFarm - Standby  **>>
 ip address 10.10.10.3 255.255.255.0
 no ip redirects
 no ip proxy-arp
 arp timeout 200
 standby 1 ip 10.10.10.1
 standby 1 timers 1 3
 standby 1 priority 50
 standby 1 preempt delay minimum 60
 standby 1 authentication <password>
end
```

## RSPAN

RSPAN allows for remote traffic monitoring in the data center. Define source and destination sessions to mirror interesting traffic to a remote VLAN captured by network analysis tools.

**Step 1** Configure a VLAN for RSPAN on the CGESM and the aggregate switch as follows:

```
(config)# vlan <vlanID>
(config-vlan)# name <vlan name>
(config-vlan)# remote-span
```

**Step 2** Create a source session as follows. This is the interface or VLAN that contains interesting traffic.

```
(config) # monitor session <session id> source vlan <VLAN IDs>
```

**Step 3** Configure the RSPAN VLAN as the target for the mirrored traffic as follows:

```
(config) # monitor session <session ID> destination remote vlan <remote vlan ID>
```

## Layer 2 Trunk Failover

The trunk failover feature may track an upstream port or a channel.

**Step 1** To assign an interface to a specific link state group, use the following command in the interface configuration sub mode:

```
(config-if)#link state group <1-2> upstream
```

**Note** Gigabit Ethernet interfaces 0/19–24 may only be configured as "upstream" devices.

**Step 2** Enable the Trunk Failover feature for the internal blade server interfaces, downstream ports, for a specific link state group.

```
interface GigabitEthernet0/1
 description blade1
 link state group <1-2> downstream

interface GigabitEthernet0/2
 description blade2
 link state group <1-2> downstream
```

**Note** Gigabit Ethernet interfaces 0/1–16 may be configured only as “downstream” devices.

**Step 3** Globally enable the trunk failover feature for a specific link state group:

```
(config)#link state track <1-2>
```

**Step 4** To validate the trunk failover configuration, use the following command:

```
show link state group detail
```